176528

STIC-Biotech/ChemLib

From: Marvich, Maria
Sent: Thursday, January 12, 2006 2:01 PM
To: STIC-Biotech/ChemLib
Subject: 09/729658

Please search SEQ ID NO:2 including interference. (391 amino acids)

Maria Bonovich Marvich, PhD
United States Patent and Trademark Office
Remsen 2B84
AU 1633
Mail Box 2C70
571-272-0774

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:_____
Date completed: 1-18-06
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#_____ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other (Specify):_____

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases: older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).
Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2006, 16:06:11 ; Search time 79 Seconds
(without alignments)
2174.647 Million cell updates/sec

Title: US-09-729-658C-2
Perfect score: 2098
Sequence: 1 MGYPEVERRELLPAAAPRER.....NMSKHTTFPGAIRLGEAPAS 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2098	100.0	391	5	ABP53104	Abp53104 Human ect
2	2098	100.0	391	5	ABB09082	Abb09082 Human EDA
3	2098	100.0	391	6	ABR42324	Abr42324 Human EDA
4	2098	100.0	391	7	ABO01928	Abo01928 Human EDA
5	2098	100.0	391	7	ADC35224	Adc35224 Human TNF
6	2098	100.0	391	7	ABW02286	Abw02286 Human EDA
7	2098	100.0	391	9	ADX06863	Adx06863 Cyclin-de
8	1873	89.3	377	5	ABB09083	Abb09083 Mouse Tab
9	1873	89.3	377	7	ABO01929	Abo01929 Mouse Ta
10	779	37.1	412	7	ADP75164	Adp75164 Fusion co
11	759	36.2	410	7	ADP75166	Adp75166 Fusion co
12	700	33.4	135	2	AAW44764	Aaw44764 Human anh
13	294	14.0	357	2	AAR95115	Aar95115 Interveni
14	294	14.0	357	2	AAW57646	Aaw57646 Collagen-
15	294	14.0	357	4	AAB64008	Aab64008 CLP-CB pr
16	294	14.0	357	4	AAB72738	Aab72738 Repetitiv
17	286	13.6	742	8	ADL91066	Adl91066 Mouse col
18	282	13.4	742	4	AAG63347	Aag63347 Amino aci
19	282	13.4	742	8	ADL91057	Adl91057 Mouse col
20	282	13.4	742	8	ADL91062	Adl91062 Mouse col
21	280.5	13.4	1838	2	AAR53257	Aar53257 Human col
22	280.5	13.4	1838	7	ADE55566	Ade55566 Human Pro
23	280.5	13.4	1838	7	ADE55570	Ade55570 Human Pro
24	280.5	13.4	1838	7	ADP65220	Adp65220 Human alp
25	280.5	13.4	1838	9	ADV70238	Adv70238 Tumor-ass
26	280.5	13.4	1838	9	ADZ09822	Adz09822 Human bre
27	280.5	13.4	1838	9	ADZ70492	Adz70492 Human pro
28	277	13.2	330	2	AAW57645	Aaw57645 Collagen-
29	277	13.2	408	2	AAW07539	Aaw07539 Collagen
30	277	13.2	408	4	AAB64007	Aab64007 CLP prote
31	277	13.2	408	4	AAB72737	Aab72737 Repetitiv
32	276	13.2	622	8	ADL91065	Adl91065 Human col
33	276	13.2	622	8	ADP56200	Adp56200 Human PRO
34	275	13.1	1466	4	AAE02534	Aae02534 Bovine al
35	275	13.1	1466	4	AAE02533	Aae02533 Bovine al
36	273.5	13.0	1014	9	ADW99577	Adw99577 Human gel
37	273	13.0	342	4	AAM39686	Aam39686 Human pol
38	273	13.0	547	3	AAY77985	Aay77985 Human col
39	273	13.0	742	4	AAG63346	Aag63346 Amino aci
40	273	13.0	742	5	ABB08642	Abb08642 Human sca
41	273	13.0	742	8	ADL91071	Adl91071 Human col
42	273	13.0	742	8	ADL91064	Adl91064 Human col
43	273	13.0	1497	6	ADA83846	Ada83846 Human COL
44	272	13.0	1497	5	ABP43711	Abp43711 Bullous p
45	268.5	12.8	342	5	ABB84830	Abb84830 Human PRO

ALIGNMENTS

RESULT 1
ABP53104
ID ABP53104 standard; protein; 391 AA.
XX
AC ABP53104;
XX
DT 11-NOV-2002 (first entry)
XX
DE Human ectodermal dysplasia protein (EDA) SEQ ID NO:6.
XX
KW Human; adipocyte complement related protein; zacrp3x2; antiinflammatory;
KW vasotropic; antimicrobial; vulnerary; antiarteriosclerotic; gene therapy;
KW adipogenesis; gluconeogenesis; glycogenolysis; lipogenesis; ischaemia;
KW glucose uptake; protein synthesis; thermogenesis; oxygen utilisation;
KW inflammation; neurotransmitter; antimicrobial; infection.
XX
OS Homo sapiens.
XX
PN WO200246417-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-US047348.
XX
PR 07-DEC-2000; 2000US-0254019P.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Haldeman BA, Thayer EC, Sheppard PO;
XX
DR WPI; 2002-657398/70.
XX
PT Novel human adipocyte complement related protein, zacrp3x2, useful for
PT modulating energy balance in mammals, and preventing injury due to
PT ischemia/inflammation.
XX
PS Disclosure; Page 111-112; 118pp; English.
XX
CC The present invention describes a human adipocyte complement related
CC protein (I) designated zacrp3x2. (I) has antiinflammatory, vasotropic,
CC antimicrobial, vulnerary and antiarteriosclerotic activities, and can be
CC used in somatic cell gene therapy, and as a modulator of
CC neurotransmission. (I) is useful for modulating energy balance in mammals
CC or for protecting endothelial cells from injury, for modulating cellular
CC metabolic reactions e.g. adipogenesis, gluconeogenesis, glycogenolysis,
CC lipogenesis, glucose uptake, protein synthesis, thermogenesis and oxygen
CC utilisation, for surgical pretreatment to prevent injury due to ischaemia

```
CC   and/or inflammation, as neurotransmitters or as modulators of
CC   neurotransmission e.g., modulates calcium ion concentration, muscle
CC   contraction, hormone separation, DNA synthesis or cell growth, inositol
CC   phosphate turnover, arachidonate release, phospholipase-C activation,
CC   gastric emptying, human neutrophil activation or superoxide anion
CC   production, for modulating nutrient uptake e.g. 2-deoxy-glucose uptake in
CC   the brain, as anti-microbial agent, as cell culture reagents in vitro
CC   studies of exogenous microorganism infection such as bacterial, viral or
CC   fungal infection, as in vivo animal models of infection, for promoting
CC   blood flow within the vasculature of a mammal by reducing the number of
CC   platelets that adhere and are activated and the size of platelet
CC   aggregates. Human zacrp3x2 is located on chromosome 5, more specifically
CC   to chromosome 5p12-p13.3. The present sequence represents human
CC   ectodermal dysplasia protein (EDA), which is given in the exemplification
CC   of the present invention
XX
SQ   Sequence 391 AA;

  Query Match             100.0%;  Score 2098;  DB 5;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-155;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              ||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 2
ABB09082
ID   ABB09082 standard; protein; 391 AA.
XX
AC   ABB09082;
XX
DT   26-JUN-2002  (first entry)
XX
DE   Human EDA1-II protein SEQ ID NO:2.
XX
KW   Human; mouse; EDA1-II; X-linked hypohidrotic ectodermal dysplasia; XLHED;
KW   DL receptor; DL gene; downless gene; ectodermal dysplasia;
KW   hair follicle growth; tooth growth; endocrine sweat gland development;
KW   mammary epithelial tissue growth; breast epithelium.
XX
OS   Homo sapiens.
XX
PN   US6355782-B1.
XX
PD   12-MAR-2002.
XX
PF   29-JUN-1999;   99US-00342681.
XX
PR   09-JUL-1998;   98US-0092279P.
PR   15-DEC-1998;   98US-0112366P.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Zonana J,  Ferguson BM,  Headon D,  Overbeek P;
XX
DR   WPI; 2002-360478/39.
DR   N-PSDB; ABL51009.
XX
PT   Novel purified hypohidrotic ectodermal dysplasias protein, useful for
PT   promoting the development of hair follicles and tooth growth.
XX
PS   Claim 1; Col 65-68; 86pp; English.
XX
CC   The present invention describes a purified hypohidrotic ectodermal
CC   dysplasias protein (I) that promotes the development of hair follicles.
CC   (I) is a promoter of hair follicle growth and tooth growth. (I) is useful
CC   for the development of hair follicles. (I) is also useful for commercial
CC   and clinical diagnostic testing for ectodermal dysplasia (ED). (I) is
CC   also useful as a stimulant for hair and tooth growth, for stimulating
CC   endocrine sweat gland development in individuals from whom the normal
CC   sweating mechanism is compromised by disease or surgery, for stimulating
CC   the growth of mammary epithelial tissue, either for reconstructing or
CC   cosmetic purposes, and for promoting or maintaining differentiation of
CC   breast epithelium. (I) is also useful for functional analysis, antibody
CC   production and patient therapy. ABL51009 to ABL51122 and ABB09082 to
CC   ABB09090 represent sequences used in the exemplification of the present
CC   invention
XX
SQ   Sequence 391 AA;

  Query Match             100.0%;  Score 2098;  DB 5;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-155;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              ||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 3
ABR42324
ID   ABR42324 standard; protein; 391 AA.
XX
AC   ABR42324;
XX
```

```
DT   11-AUG-2003  (first entry)
XX
DE   Human EDA protein.
XX
KW   Human; EDA; tumour necrosis factor; ligand; cytostatic; immunomodulator;
KW   osteopathic.
XX
OS   Homo sapiens.
XX
PN   WO2003040307-A2.
XX
PD   15-MAY-2003.
XX
PF   25-JUL-2002; 2002WO-US023782.
XX
PR   27-JUL-2001; 2001US-0307838P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Hilbert DH,  Rosen CA;
XX
DR   WPI; 2003-430659/40.
DR   N-PSDB; ACC57910.
XX
PT   New heteromultimeric complex having a first polypeptide member of the
PT   tumor necrosis factor (TNF) ligand family, and a second different member
PT   of TNF ligand family, useful for treating cancer, osteoporosis or an
PT   autoimmune disease.
XX
PS   Disclosure; Page 387-388; 388pp; English.
XX
CC   The present sequence is the protein sequence of human EDA polypeptide.
CC   The invention relates to compositions comprising heterotrimeric complexes
CC   of tumour necrosis factor (TNF) ligand family members, and their use in
CC   the detection, prevention and treatment of disease. In preferred
CC   embodiments, the heterotrimeric complex comprises full-length or
CC   extracellular portions of EDA together with full-length or extracellular
CC   portions of other TNF ligand family members. The heterotrimeric complexes
CC   of the invention are useful for treating an autoimmune disease, cancer or
CC   osteoporosis, and particularly for inhibiting cancer cell proliferation,
CC   increasing B cell proliferation, or inducing apoptosis of T cells
XX
SQ   Sequence 391 AA;

  Query Match             100.0%;  Score 2098;  DB 6;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-155;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              ||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 4
ABO01928
ID   ABO01928 standard; protein; 391 AA.
XX
AC   ABO01928;
XX
DT   15-AUG-2003  (first entry)
XX
DE   Human EDA1-II protein.
XX
KW   Human; EDA1-II; hair follicle development; tooth development;
KW   sweat gland development; ectodermal disease; alopecia; skin trauma; burn;
KW   X-linked hypohidrotic ectodermal dysplasia; XLHED; breast cancer; HED;
KW   autosomal recessive hypohidrotic ectodermal dysplasia; hirsutism;
KW   ectopic teeth.
XX
OS   Homo sapiens.
XX
PN   US2003023991-A1.
XX
PD   30-JAN-2003.
XX
PF   04-DEC-2000; 2000US-00729658.
XX
PR   09-JUL-1998;   98US-0092279P.
PR   15-DEC-1998;   98US-0112366P.
PR   29-JUN-1999;   99US-00342681.
XX
PA   (ZONA/) ZONANA J.
PA   (FERG/) FERGUSON B M.
PA   (HEAD/) HEADON D.
PA   (OVER/) OVERBEEK P.
XX
PI   Zonana J,  Ferguson BM,  Headon D,  Overbeek P;
XX
DR   WPI; 2003-456312/43.
DR   N-PSDB; ACD07895.
XX
PT   Increasing or decreasing hair follicle development, tooth development, or
PT   sweat gland development, in a tissue, involves modulating ectodermal
PT   dysplasia protein isoform II activity in the tissue.
XX
PS   Claim 26; Fig 1; 102pp; English.
XX
CC   The invention relates to a method of increasing or decreasing one or more
CC   of hair follicle development, tooth development, or sweat gland
CC   development, in a tissue, involves altering ectodermal dysplasia (EDA1)
CC   isoform II (EDA1-II) activity in the tissue. The method is useful for
CC   increasing or decreasing hair follicle development, tooth development or
CC   sweat gland development, in a tissue of a patient suffering from an
CC   ectodermal disease such as X-linked hypohidrotic ectodermal dysplasia
CC   (XLHED), autosomal recessive hypohidrotic ectodermal dysplasia (HED) or
CC   alopecia. The method is useful for decreasing hair follicle development
CC   in a subject suffering from hirsutism. The method is also useful for
CC   treating breast cancer and ectopic teeth, burns and trauma of skin due to
CC   surgery. The present sequence represents the amino acid sequence of human
CC   EDA1-II
XX
SQ   Sequence 391 AA;

  Query Match             100.0%;  Score 2098;  DB 7;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-155;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
                 ||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
```

```
RESULT 5
ADC35224
ID   ADC35224 standard; protein; 391 AA.
XX
AC   ADC35224;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human TNF ligand family member #21.
XX
KW   human; tumour necrosis factor; TNF ligand; endokine alpha;
KW   excessive bone resorption disorder; osteoporosis; Paget's disease;
KW   arterial calcification.
XX
OS   Homo sapiens.
XX
PN   US2003100074-A1.
XX
PD   29-MAY-2003.
XX
PF   15-AUG-2002; 2002US-00218547.
XX
PR   16-AUG-2001; 2001US-0312542P.
PR   30-OCT-2001; 2001US-0330761P.
XX
PA   (YUGG/) YU G.
PA   (NIJJ/) NI J.
PA   (ROSE/) ROSEN C A.
PA   (NARD/) NARDELLI B.
XX
PI   Yu G,  Ni J,  Rosen CA,  Nardelli B;
XX
DR   WPI; 2003-696072/66.
DR   N-PSDB; ADC35223.
XX
PT   New Endokine alpha gene useful for preparing a composition for treating a
PT   disease associated with excessive or insufficient bone resorption e.g.,
PT   osteoporosis, Paget's disease or arterial calcification.
XX
PS   Disclosure; SEQ ID NO 42; 145pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule encoding a
CC   tumour necrosis factor family ligand. A composition comprising the
CC   isolated antibody or its fragment is used for treating an individual in
CC   need of decreased level of endokine alpha activity. The endokine alpha
CC   polypeptide present in a heterotrimeric complex is used for treating an
CC   individual having a disorder associated with excessive bone resorption,
CC   e.g. osteoporosis, Paget's disease or arterial calcification. Treating an
CC   individual having a disorder associated with insufficient bone resorption
CC   comprises administering an endokine alpha antagonist, which is the
CC   antibody that binds specifically to endokine alpha polypeptide. The
CC   present sequence represents the amino acid sequence of a tumour necrosis
CC   factor family ligand.
XX
SQ   Sequence 391 AA;

  Query Match             100.0%;  Score 2098;  DB 7;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-155;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
                 ||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
```

```
RESULT 6
ABW02286
ID   ABW02286 standard; protein; 391 AA.
XX
AC   ABW02286;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human EDA protein.
XX
KW   Gastrointestinal tract disorder; tumour necrosis factor; TNF; DR3; TR6;
KW   TNF-gamma-beta protein; inflammatory bowel disease; Crohn's disease;
KW   ulcerative colitis; EDA protein; human.
XX
OS   Homo sapiens.
XX
PN   US2003198640-A1.
XX
PD   23-OCT-2003.
XX
PF   06-DEC-2002; 2002US-00310793.
XX
PR   07-NOV-1994;   94WO-US012880.
PR   05-JUN-1995;   95US-00461246.
PR   09-JAN-1998;   98US-00005020.
PR   09-FEB-1998;   98US-0074047P.
PR   07-AUG-1998;   98US-00131237.
PR   08-FEB-1999;   99US-00246129.
PR   30-APR-1999;   99US-0131963P.
PR   03-MAY-1999;   99US-0132227P.
PR   13-MAY-1999;   99US-0134067P.
PR   08-FEB-2000; 2000US-0180908P.
PR   27-APR-2000; 2000US-00559290.
```

PR 07-JUL-2000; 2000US-0216879P.
PR 26-MAR-2001; 2001US-0278449P.
PR 06-JUL-2001; 2001US-00899059.
PR 24-AUG-2001; 2001US-0314381P.
PR 07-DEC-2001; 2001US-0336695P.
PR 23-AUG-2002; 2002US-00226294.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Yu G, Ni J, Rosen CA, Zhang J, Wei P;
XX
DR WPI; 2003-852773/79.
DR N-PSDB; AAD63922.
XX
PT Use of tumor necrosis factor gamma-beta antagonists for treating or
PT ameliorating a disease or disorders of the gastrointestinal tract, e.g.
PT inflammatory bowel disease, Crohn's disease or ulcerative colitis.
XX
PS Disclosure; Page 156-157; 0pp; English.
XX
CC The invention relates to methods and compositions for treating or
CC ameliorating a disease or disorder of the gastrointestinal tract. The
CC method involves administering a composition comprising tumour necrosis
CC factor (TNF)-gamma-beta and its receptors DR3 and TR6 to a person with,
CC or suspected of having the disease or disorder. The antagonist of TNF-
CC gamma-beta is useful for treating or ameliorating a gastrointestinal
CC tract disease or disorder, specifically an inflammatory bowel disease,
CC e.g. Crohn's disease or ulcerative colitis. The present sequence is human
CC EDA protein. This sequence is used to illustrate the method of the
CC invention
XX
SQ Sequence 391 AA;

Query Match 100.0%; Score 2098; DB 7; Length 391;
Best Local Similarity 100.0%; Pred. No. 5.9e-155;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              ||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
```

RESULT 7
ADX06863
ID ADX06863 standard; protein; 391 AA.
XX
AC ADX06863;
XX
DT 21-APR-2005 (first entry)
XX
DE Cyclin-dependent kinase modulation biomarker SEQ ID NO 1428.
XX
KW cytostatic; cyclin-dependent kinase; cdk; biomarker.
XX
OS Homo sapiens.
XX
PN WO2005012875-A2.
XX
PD 10-FEB-2005.
XX
PF 29-JUL-2004; 2004WO-US024424.
XX
PR 29-JUL-2003; 2003US-0490890P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Li M, Rupnow BA, Webster KR, Jackson DG, Wong TW;
XX
DR WPI; 2005-163068/17.
DR N-PSDB; ADX06862.
XX
PT Biomarkers useful for predicting or determining the response of a mammal
PT to a cancer treatment comprising administration of a modulator of cyclin-
PT dependent kinase activity.
XX
PS Claim 5; SEQ ID NO 1428; 141pp; English.
XX
CC This invention describes a novel method of predicting or determining
CC whether a mammal will respond or is responding to an anti-cancer agent
CC that modulates cyclin-dependent kinase (cdk) activity. The method
CC comprises measuring the level of one or more biomarkers selected from
CC 2774 biomarkers given in the specification (nucleotide sequence SEQ ID
CC NO:1246 (Genbank EST W28729) is especially preferred). The method of the
CC invention is utilized in a kit for determining or predicting whether
CC patient would be susceptible or resistant to treatment by an agent
CC modulating cdk activity. The invention also describes a method for
CC utilizing individualized genetic profiles for treating diseases and
CC disorders based on patient's response and molecular level, specialized
CC microarrays comprising the biomarkers described, antibodies directed
CC against the biomarkers and a cell culture model to identify biomarkers.
CC The cdk modulator is preferably N-5-[[5-(1,1-Dimethylethyl)-2-
CC oxazolyl]methyl]thio]-2-thiazolyl-4-piperidine carboxamide, 0.5-L-
CC tartaric acid salt. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. This
CC sequence represents a biomarker used in the method of the invention.
XX
SQ Sequence 391 AA;

Query Match 100.0%; Score 2098; DB 9; Length 391;
Best Local Similarity 100.0%; Pred. No. 5.9e-155;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
```

```
Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 8
ABB09083
ID   ABB09083 standard; protein; 377 AA.
XX
AC   ABB09083;
XX
DT   26-JUN-2002  (first entry)
XX
DE   Mouse Tabby (Ta) protein SEQ ID NO:4.
XX
KW   Human; mouse; EDA1-II; X-linked hypohidrotic ectodermal dysplasia; XLHED;
KW   DL receptor; DL gene; downless gene; ectodermal dysplasia;
KW   hair follicle growth; tooth growth; endocrine sweat gland development;
KW   mammary epithelial tissue growth; breast epithelium.
XX
OS   Mus musculus.
XX
PN   US6355782-B1.
XX
PD   12-MAR-2002.
XX
PF   29-JUN-1999;   99US-00342681.
XX
PR   09-JUL-1998;   98US-0092279P.
PR   15-DEC-1998;   98US-0112366P.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Zonana J,  Ferguson BM,  Headon D,  Overbeek P;
XX
DR   WPI; 2002-360478/39.
DR   N-PSDB; ABL51010.
XX
PT   Novel purified hypohidrotic ectodermal dysplasias protein, useful for
PT   promoting the development of hair follicles and tooth growth.
XX
PS   Example 1; Col 71-74; 86pp; English.
XX
CC   The present invention describes a purified hypohidrotic ectodermal
CC   dysplasias protein (I) that promotes the development of hair follicles.
CC   (I) is a promoter of hair follicle growth and tooth growth. (I) is useful
CC   for the development of hair follicles. (I) is also useful for commercial
CC   and clinical diagnostic testing for ectodermal dysplasia (ED). (I) is
CC   also useful as a stimulant for hair and tooth growth, for stimulating
CC   endocrine sweat gland development in individuals from whom the normal
CC   sweating mechanism is compromised by disease or surgery, for stimulating
CC   the growth of mammary epithelial tissue, either for reconstructing or
CC   cosmetic purposes, and for promoting or maintaining differentiation of
CC   breast epithelium. (I) is also useful for functional analysis, antibody
CC   production and patient therapy. ABL51009 to ABL51122 and ABB09082 to
CC   ABB09090 represent sequences used in the exemplification of the present
CC   invention
XX
SQ   Sequence 377 AA;

  Query Match             89.3%;  Score 1873;  DB 5;  Length 377;
  Best Local Similarity   90.8%;  Pred. No. 1.9e-137;
  Matches  355;  Conservative    2;  Mismatches   20;  Indels   14;  Gaps    1;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| ||||||||||||||| |||||||||| ||||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||| |||||| | ||||||||| | |||  ||| ||||||  :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                |  :||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db        121 DPLPPESQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGEL------- 293

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 -------EVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 346

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        347 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 377


RESULT 9
ABO01929
ID   ABO01929 standard; protein; 377 AA.
XX
AC   ABO01929;
XX
DT   15-AUG-2003  (first entry)
XX
DE   Mouse Ta protein.
XX
KW   Mouse; Tabby; hair follicle development; tooth development; Ta;
KW   sweat gland development; ectodermal disease; alopecia; skin trauma; burn;
KW   X-linked hypohidrotic ectodermal dysplasia; XLHED; breast cancer; HED;
KW   autosomal recessive hypohidrotic ectodermal dysplasia; hirsutism;
KW   ectopic teeth.
XX
OS   Mus musculus.
XX
PN   US2003023991-A1.
XX
PD   30-JAN-2003.
XX
PF   04-DEC-2000; 2000US-00729658.
XX
PR   09-JUL-1998;   98US-0092279P.
PR   15-DEC-1998;   98US-0112366P.
PR   29-JUN-1999;   99US-00342681.
XX
PA   (ZONA/) ZONANA J.
PA   (FERG/) FERGUSON B M.
PA   (HEAD/) HEADON D.
PA   (OVER/) OVERBEEK P.
XX
PI   Zonana J,  Ferguson BM,  Headon D,  Overbeek P;
XX
DR   WPI; 2003-456312/43.
DR   N-PSDB; ACD07896.
XX
PT   Increasing or decreasing hair follicle development, tooth development, or
PT   sweat gland development, in a tissue, involves modulating ectodermal
PT   dysplasia protein isoform II activity in the tissue.
XX
PS   Example 1; Fig 1; 102pp; English.
XX
CC   The invention relates to a method of increasing or decreasing one or more
CC   of hair follicle development, tooth development, or sweat gland
CC   development, in a tissue, involves altering ectodermal dysplasia (EDA1)
```

```
CC   isoform II (EDA1-II) activity in the tissue. The method is useful for
CC   increasing or decreasing hair follicle development, tooth development or
CC   sweat gland development, in a tissue of a patient suffering from an
CC   ectodermal disease such as X-linked hypohidrotic ectodermal dysplasia
CC   (XLHED), autosomal recessive hypohidrotic ectodermal dysplasia (HED) or
CC   alopecia. The method is useful for decreasing hair follicle development
CC   in a subject suffering from hirsutism. The method is also useful for
CC   treating breast cancer and ectopic teeth, burns and trauma of skin due to
CC   surgery. The present sequence represents the amino acid sequence of mouse
CC   Tabby (Ta)
XX
SQ   Sequence 377 AA;

  Query Match             89.3%;  Score 1873;  DB 7;  Length 377;
  Best Local Similarity   90.8%;  Pred. No. 1.9e-137;
  Matches  355;  Conservative    2;  Mismatches   20;  Indels   14;  Gaps    1;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| |||||||||||||| |||||||||| ||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||| |||||| | ||||||||| | |||  ||| |||||| :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                | :||| |||||||||||||| |||||||||||||||| ||||||||||||||||
Db        121 DPLPPESQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGEL------- 293

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 -------EVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 346

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        347 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 377


RESULT 10
ADP75164
ID   ADP75164 standard; protein; 412 AA.
XX
AC   ADP75164;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Fusion construct ps1236 Fc:EDA1 protein.
XX
KW   fusion protein; immunoglobulin; Fc region; TNF; tumour necrosis factor;
KW   FasL; TNF-alpha; TNF-gamma; TRAIL; EDA1; EDA2; dermatological;
KW   depilatory; gene therapy; cellular therapy; ectodermal dysplasia;
KW   X-linked immune deficiency; hyper IgM; alopaecia; hirsutism; sweat gland;
KW   sebaceous gland.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .15
FT                   /note= "HA signal"
FT   Region          18. .243
FT                   /note= "hIgG Fc aa248-473 of gi2765420"
FT   Peptide         246. .259
FT                   /note= "linker peptide"
FT   Region          266. .412
FT                   /note= "hED1 aa245-391"
XX
PN   DE10205583-A1.
XX
PD   21-AUG-2003.
XX
PF   11-FEB-2002; 2002DE-01005583.
XX
PR   10-FEB-2002; 2002DE-01005368.
XX
PA   (APOX-) APOXIS SA.
XX
PI   Gaide O,  Schneider P,  Tschopp J;
XX
DR   WPI; 2003-647319/62.
DR   N-PSDB; ADP75163.
XX
PT   New recombinant fusion protein, useful for treating deficiency of tumor
PT   necrosis factor ligand, particularly ectodermal dysplasia, comprises the
PT   ligand and immunoglobulin fragment.
XX
PS   Claim 8; Fig D; 22pp; German.
XX
CC   This invention describes a novel recombinant fusion protein which
CC   comprises an immunoglobulin Fc segment, the extracellular region of a
CC   tumour necrosis factor (TNF) and optionally a transition region,
CC   including a linker. The TNF ligands used in the method include FasL, TNF-
CC   alpha or TNF-gamma, TRAIL or most preferred, EDA1 or 2. The Fc fragment
CC   contains the hinge region and domains CH2 and CH3 of the Fc segment of an
CC   IgG, particularly human. The linker optionally includes a protease
CC   cleavage site. The N-terminus of construct may have a signal sequence,
CC   secretory signal sequence and/or tag (e.g. Flag or His). The products of
CC   the invention have dermatological and depilatory activity. The construct,
CC   or other therapeutic agents, are administered parenterally (especially
CC   intravenously or intra-arterially) during pregnancy, especially as early
CC   as possible, e.g. for humans, in the first 3 months, every second day for
CC   at least 14 days. Nucleic acid encoding the construct, related vectors
CC   and transfected host cells, are useful in gene/cellular therapy of
CC   genetic disorders associated with defective expression of TNF ligand,
CC   particularly ectodermal dysplasia (specifically the X-linked hypohydrotic
CC   form or X-linked immune deficiency with hyper IgM) and also alopaecia,
CC   hirsutism and inadequate functioning of the sweat and sebaceous glands.
CC   The fusion protein construct is soluble but still retains ability to
CC   mimic the physiological effects of TNF ligand. This sequence represents
CC   the fusion construct Fc-EDA1.
XX
SQ   Sequence 412 AA;

  Query Match             37.1%;  Score 779;  DB 7;  Length 412;
  Best Local Similarity   87.6%;  Pred. No. 3.8e-52;
  Matches  155;  Conservative    1;  Mismatches   15;  Indels    6;  Gaps    2;

Qy        215 GPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 274
              |   |   | ||  |    | |:       ||||||||||||||||||||||||||||
Db        242 GKRSPQPQPKPQPKP---EPEGSLQ---VDENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 295

Qy        275 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTR 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTR 355

Qy        335 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 412


RESULT 11
ADP75166
ID   ADP75166 standard; protein; 410 AA.
XX
AC   ADP75166;
XX
DT   12-AUG-2004  (first entry)
XX
```

```
DE   Fusion construct ps1236 Fc:EDA2 protein.
XX
KW   fusion protein; immunoglobulin; Fc region; TNF; tumour necrosis factor;
KW   FasL; TNF-alpha; TNF-gamma; TRAIL; EDA1; EDA2; dermatological;
KW   depilatory; gene therapy; cellular therapy; ectodermal dysplasia;
KW   X-linked immune deficiency; hyper IgM; alopaecia; hirsutism; sweat gland;
KW   sebaceous gland.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .15
FT                   /note= "HA signal"
FT   Region          18. .243
FT                   /note= "hIgG1 Fc aa248-473 of gi2765420"
FT   Peptide         246. .259
FT                   /note= "linker peptide"
FT   Region          266. .410
FT                   /note= "hEDA2 aa245-389"
XX
PN   DE10205583-A1.
XX
PD   21-AUG-2003.
XX
PF   11-FEB-2002; 2002DE-01005583.
XX
PR   10-FEB-2002; 2002DE-01005368.
XX
PA   (APOX-) APOXIS SA.
XX
PI   Gaide O,  Schneider P,  Tschopp J;
XX
DR   WPI; 2003-647319/62.
DR   N-PSDB; ADP75165.
XX
PT   New recombinant fusion protein, useful for treating deficiency of tumor
PT   necrosis factor ligand, particularly ectodermal dysplasia, comprises the
PT   ligand and immunoglobulin fragment.
XX
PS   Claim 8; Fig E; 22pp; German.
XX
CC   This invention describes a novel recombinant fusion protein which
CC   comprises an immunoglobulin Fc segment, the extracellular region of a
CC   tumour necrosis factor (TNF) and optionally a transition region,
CC   including a linker. The TNF ligands used in the method include FasL, TNF-
CC   alpha or TNF-gamma, TRAIL or most preferred, EDA1 or 2. The Fc fragment
CC   contains the hinge region and domains CH2 and CH3 of the Fc segment of an
CC   IgG, particularly human. The linker optionally includes a protease
CC   cleavage site. The N-terminus of construct may have a signal sequence,
CC   secretory signal sequence and/or tag (e.g. Flag or His). The products of
CC   the invention have dermatological and depilatory activity. The construct,
CC   or other therapeutic agents, are administered parenterally (especially
CC   intravenously or intra-arterially) during pregnancy, especially as early
CC   as possible, e.g. for humans, in the first 3 months, every second day for
CC   at least 14 days. Nucleic acid encoding the construct, related vectors
CC   and transfected host cells, are useful in gene/cellular therapy of
CC   genetic disorders associated with defective expression of TNF ligand,
CC   particularly ectodermal dysplasia (specifically the X-linked hypohydrotic
CC   form or X-linked immune deficiency with hyper IgM) and also alopaecia,
CC   hirsutism and inadequate functioning of the sweat and sebaceous glands.
CC   The fusion protein construct is soluble but still retains ability to
CC   mimic the physiological effects of TNF ligand. This sequence represents
CC   the fusion construct Fc-EDA2.
XX
SQ   Sequence 410 AA;

  Query Match             36.2%;  Score 759;  DB 7;  Length 410;
  Best Local Similarity   86.4%;  Pred. No. 1.4e-50;
  Matches  153;  Conservative    1;  Mismatches   15;  Indels    8;  Gaps    3;

Qy        215 GPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 274
              |   |   | ||  |    | |:      ||||||||||||||||||||||||||||||

Db        242 GKRSPQPQPKPQPKP---EPEGSLQ---VDENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 295

Qy        275 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTR 334
              |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||

Db        296 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQ--VYYINFTDFASYEVVVDEKPFLQCTR 353

Qy        335 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        354 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 410


RESULT 12
AAW44764
ID   AAW44764 standard; protein; 135 AA.
XX
AC   AAW44764;
XX
DT   01-JUN-1998  (first entry)
XX
DE   Human anhidrotic ectodermal dysplasia protein.
XX
KW   Human; anhidrotic ectodermal dysplasia; X chromosome; genetic linkage;
KW   translocation; CpG island; foetal development; hair; sweat gland; tooth.
XX
OS   Homo sapiens.
XX
PN   US5700926-A.
XX
PD   23-DEC-1997.
XX
PF   22-JUL-1996;   96US-00684672.
XX
PR   27-APR-1993;   93US-00052997.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   De La Chapelle A,  Srivastava AK,  Kere J,  Schlessinger D;
XX
DR   WPI; 1998-062436/06.
DR   N-PSDB; AAV05851.
XX
PT   Human anhidrotic ectodermal dysplasia gene - useful for research into
PT   hair growth.
XX
PS   Disclosure; Fig 7; 37pp; English.
XX
CC   This is the amino acid sequence of the protein encoded by the human
CC   anhidrotic ectodermal dysplasia (EDA) gene. The gene has been mapped to
CC   the region Xq12-q13 by genetic linkage analysis and has been shown to
CC   contain a 200 kb intron inserted in the 3' end of the coding sequence.
CC   Deficiencies in the gene are observed by translocations with a breakpoint
CC   in the transcribed CpG island 3 at the Xq12-q13 locus. The EDA gene can
CC   be used to study the dynamics of EDA gene expression during foetal
CC   development, and processes affecting normal hair growth in adults. The
CC   EDA gene can also be used to study hair, sweat gland and tooth formation
CC   and growth, and ectodermal dysplasias
XX
SQ   Sequence 135 AA;

  Query Match             33.4%;  Score 700;  DB 2;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-46;
  Matches  132;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQ 132
              ||||||||||||
```

```
Db        121 AALHSDSQDGHQ 132


RESULT 13
AAR95115
ID   AAR95115 standard; protein; 357 AA.
XX
AC   AAR95115;
XX
DT   25-MAR-2003  (revised)
DT   03-FEB-1997  (first entry)
XX
DE   Intervening sequence contg. collagen like protein (CLP)-CB.
XX
KW   Polymer; repeat unit; natural collagen; intervening oligopeptide; fibre;
KW   film; membrane; emulsion; coating; collagen like protein;
KW   specific binding material; catalyst; purification agent; composite;
KW   laminate; adhesive; cell growth surface; affinity column;
KW   biological material support; wound dressing; in vivo prothesis.
XX
OS   Synthetic.
XX
PN   US5514581-A.
XX
PD   07-MAY-1996.
XX
PF   06-NOV-1990;   90US-00609716.
XX
PR   04-NOV-1986;   86US-00927258.
PR   29-OCT-1987;   87US-00114618.
PR   09-NOV-1988;   88US-00269429.
PR   07-NOV-1989;   89WO-US005016.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J,  Ferrari FA;
XX
DR   WPI; 1996-238772/24.
XX
PT   DNA encoding protein contg. repeated fibroin derived segments - linked by
PT   oligopeptide with cell adhesion properties useful, e.g. in wound
PT   dressings.
XX
PS   Example 3; Col 101-104; 71pp; English.
XX
CC   Novel DNA sequence encodes a polymer comprising segments of repeating
CC   units of 3-9 amino acids from natural collagen, i.e. the present
CC   intervening sequence contg. collagen like protein (CLP)-CB, able to
CC   assemble into aligned structures formable into articles. The polymer
CC   comprises at least 2 segments joined by an unaligned intervening
CC   oligopeptide, other than the repeating unit. The polymer can be used to
CC   make fibres, films, membranes, emulsions, coatings, etc., useful as, e.g.
CC   specific binding materials, catalysts, purifican. agents, composites,
CC   laminates, adhesives, cell growth surfaces, affinity columns and supports
CC   for biological materials. Typical applications include wound dressings,
CC   and in vivo protheses. The polymer produces articles with good mechanical
CC   properties, and the intervening oligopeptide can provide a ligand for
CC   binding a mol., antibody, etc., or a chemically reactive site for
CC   coupling to proteins, etc. . (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 357 AA;

  Query Match             14.0%;  Score 294;  DB 2;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 2e-14;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy         71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              :||      |  |:|| :| : | |   |  |        | | |   |  ||   |
Db         71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGAPGPPGPPGPPGPPGLPGPKG 130

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
              |                          | :   |  :|: ||          ||||| |||
Db        131 D------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy        189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
              |||| ||| |||| || ||:||
Db        166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy        213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
               | ||||||||| ||||| || |    || :
Db        226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258


RESULT 14
AAW57646
ID   AAW57646 standard; peptide; 357 AA.
XX
AC   AAW57646;
XX
DT   27-AUG-1998  (first entry)
XX
DE   Collagen-like polymer.
XX
KW   Collagen-like polymer; synthetic polymer; fibre coating;
KW   prosthetic device; catalytic substance.
XX
OS   Synthetic.
XX
PN   US5773249-A.
XX
PD   30-JUN-1998.
XX
PF   02-MAY-1996;   96US-00642255.
XX
PR   04-NOV-1986;   86US-00927258.
PR   29-OCT-1987;   87US-00114618.
PR   09-NOV-1988;   88US-00269429.
PR   06-NOV-1990;   90US-00609716.
PR   12-NOV-1991;   91US-00791960.
PR   05-NOV-1992;   92US-00972032.
PR   22-DEC-1995;   95US-00577046.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Ferrari FA,  Cappello J;
XX
DR   WPI; 1998-387004/33.
XX
PT   Recombinant collagen-like polymers - useful for making gels, films,
PT   fibres, etc.
XX
PS   Example 2; Col 23; 93pp; English.
XX
CC   This sequence represents a unnatural collagen-like polymer of the
CC   invention. The products may be used as films, fibres, moulded objects and
CC   admixed with other natural or synthetic polymers or coatings on fibres,
CC   films, labware or other surfaces, e.g. prosthetic devices. The polymers
CC   may be used for binding a wide variety of specific binding materials, as
CC   catalytic substances (where the amino acid sequence may specifically
CC   chelate a wide variety of elements), as purification media, composites,
CC   laminates or adhesives. They may also be combined with inorganic or
CC   organic materials such as carbon fibres, nylon fibres, nitrocellulose,
CC   etc., as flask coatings or in synthetic matrices for the growth and study
CC   of cells, as affinity columns or as supports for biological materials.
CC   The polymers have collagen-like properties, but may be easily expressed
CC   in micro-organisms in high efficiency. The new sequences can be tailored
CC   to give the desired properties
XX
SQ   Sequence 357 AA;

  Query Match             14.0%;  Score 294;  DB 2;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 2e-14;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy         71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              :||      |  |:|| :| : | |   |  |        | | |   |  ||   |
```

```
Db         71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGPLPGPKG 130

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
              |                          | :   |  :|: ||          |||||  |||
Db        131 D------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy        189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
              ||||  |||  ||||  ||  ||:||
Db        166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy        213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
               | |||||||||  |||||  || |    || :
Db        226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258
```

```
RESULT 15
AAB64008
ID   AAB64008 standard; protein; 357 AA.
XX
AC   AAB64008;
XX
DT   19-MAR-2001  (first entry)
XX
DE   CLP-CB protein sequence SEQ ID 66.
XX
KW   Proteinaceous polymer; repeat unit; structural polymer; coating; film;
KW   fibre; membrane; adhesive; emulsion; laminate; keratin; collagen.
XX
OS   Synthetic.
XX
PN   US6140072-A.
XX
PD   31-OCT-2000.
XX
PF   07-JUN-1995;   95US-00475411.
XX
PR   04-NOV-1986;   86US-00927258.
PR   29-OCT-1987;   87US-00114618.
PR   09-NOV-1988;   88US-00269429.
PR   06-NOV-1990;   90US-00609716.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J,  Ferrari FA;
XX
DR   WPI; 2001-048958/06.
XX
PT   New DNA encoding a polymer with strands of repeating units of natural
PT   protein joined by intervening oligopeptide for producing high molecular
PT   weight polymers of amino acids.
XX
PS   Example 2; Col 41; 73pp; English.
XX
CC   This invention relates to DNA encoding a proteinaceous polymer. The
CC   polymer comprises strands of repeating units of a natural protein capable
CC   of assembling into aligned structures, with at least 2 strands joined by
CC   an intervening oligopeptide other than the repeating units. The
CC   intervening oligopeptide is unaligned and the polymer has individual
CC   strands of the same or different repeating units. The DNA is useful for
CC   producing high molecular weight polymers of amino acids based on
CC   biologically and chemically active structural polymers. These polymers
CC   may be used to provide a variety of structures for different purposes,
CC   and to produce articles including coatings, or other (non)structural
CC   components, e.g. fibres, films, membranes, adhesives or emulsions, or
CC   with other compounds and/or compositions to form composites or laminates.
CC   Peptide sequences AAB63971-B63991 represent monomer sequences which can
CC   be used in the polymers of the invention. Oligonucleotide sequences
CC   AAF23370 - AAF23386 and amino acid sequences AAB63992 - AAB64002 are used
CC   in the construction of SLP and FCB-SLP polymers. Oligonucleotide
CC   sequences AAF23387 - AAF23397 and amino acid sequences AAB64003 -
CC   AAB64008 are used in the construction of CLP (collagen like protein)
CC   polymers. Oligonucleotide sequences AAF23398 - AAF23409 and amino acid
CC   sequences AAB64009 - AAB64014 are used in the construction of keratin
CC   polymers. Proteins and peptides represented by sequences AAB64015 -
CC   AAB64049 are examples of polymers of the invention
XX
SQ   Sequence 357 AA;

  Query Match             14.0%;  Score 294;  DB 4;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 2e-14;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy         71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              :||     |  |:|| :| : | |   | |      | | |   |  ||   |
Db         71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGLPGPKG 130

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
              |                          | :   |  :|: ||          |||||  |||
Db        131 D------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy        189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
              ||||  |||  ||||  ||  ||:||
Db        166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy        213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
               | |||||||||  |||||  || |    || :
Db        226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258
```

```
Search completed: January 13, 2006, 16:13:31
Job time : 81 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2006, 16:08:46 ; Search time 16 Seconds
(without alignments)
2351.297 Million cell updates/sec

Title: US-09-729-658C-2
Perfect score: 2098
Sequence: 1 MGYPEVERRELLPAAAPRER.....NMSKHTTFFGAIRLGEAPAS 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	280.5	13.4	1838	1	CGHU1V	collagen alpha 1(V
2	277	13.2	1464	2	S59856	collagen alpha 1(I
3	275.5	13.1	1433	2	A46053	bullous pemphigoid
4	275	13.1	1049	1	CGBO7S	collagen alpha 1(I
5	273.5	13.0	1843	2	S18803	collagen alpha 1(V
6	273	13.0	1532	2	A61262	collagen alpha 1(X
7	268.5	12.8	742	2	JC7595	scavenger receptor
8	267.5	12.8	316	2	T20497	hypothetical prote
9	267	12.7	886	2	I50694	collagen alpha 1(I
10	267	12.7	1315	2	A56101	collagen alpha 1(X
11	267	12.7	1774	2	B56101	collagen alpha 1(X
12	266	12.7	779	1	CGBO1S	collagen alpha 1(I
13	265.5	12.7	1464	1	CGHU1S	collagen alpha 1(I
14	265	12.6	1763	2	S16366	collagen alpha 2(I
15	264	12.6	1690	1	CGHU1B	collagen alpha 4(I
16	263.5	12.6	1146	2	A38587	collagen, cornea-s
17	263.5	12.6	1419	2	A41182	collagen alpha 1(I
18	263.5	12.6	1487	2	B41182	collagen alpha 1(I
19	263	12.5	1366	1	CGHU2S	collagen alpha 2(I
20	260.5	12.4	675	2	S20819	collagen alpha 3(I
21	259.5	12.4	302	2	T15936	hypothetical prote
22	259.5	12.4	674	2	S23297	collagen alpha 1(X
23	258.5	12.3	1373	1	A43291	collagen alpha 2(I
24	258	12.3	674	2	S13301	collagen alpha 1(X
25	258	12.3	680	1	CGHU1D	collagen alpha 1(X
26	258	12.3	1142	2	JX0369	collagen alpha 1(X
27	257.5	12.3	312	2	T25048	hypothetical prote
28	257.5	12.3	1453	2	S21626	collagen alpha 1(I
29	256.5	12.2	688	2	A53330	collagen alpha 2(I
30	256.5	12.2	1487	1	CGHU6C	collagen alpha 1(I
31	256	12.2	1466	1	CGHU7L	collagen alpha 1(I
32	254	12.1	1670	1	CGHU3B	collagen alpha 3(I
33	253.5	12.1	964	1	CGCH2S	collagen alpha 2(I
34	253.5	12.1	1414	1	S23809	collagen alpha 2(I
35	253	12.1	744	2	S15435	collagen alpha 1(V
36	252.5	12.0	1388	2	A53317	collagen alpha 1(X
37	252	12.0	1418	2	T45467	collagen alpha 1(I
38	251.5	12.0	684	2	A53019	collagen alpha 1(X
39	251	12.0	1042	1	CGCH1S	collagen alpha 1(I
40	251	12.0	1806	1	CGHU1E	collagen alpha 1(X
41	250.5	11.9	301	2	T21314	hypothetical prote
42	250.5	11.9	329	2	T32783	hypothetical prote
43	250	11.9	102	2	B34770	ORF2 protein - sai
44	250	11.9	471	2	A39024	collagen alpha 3(I
45	250	11.9	633	2	B40983	collagen alpha 1(X

ALIGNMENTS

RESULT 1
CGHU1V
collagen alpha 1(V) chain precursor - human
N;Alternate names: procollagen alpha 1(V) chain
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004
C;Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
J. Biol. Chem. 266, 24727-24733, 1991
A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of
A;Reference number: S18802; MUID:92105142; PMID:1722213
A;Accession: S18802
A;Molecule type: mRNA
A;Residues: 1-1838 <GRE>
A;Cross-references: UNIPROT:Q15094; UNIPARC:UPI000006E982; GB:M76729; NID:g189519; PIDN
R;Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.
J. Biol. Chem. 266, 13124-13129, 1991
A;Title: Complete primary structure of human collagen alpha-1(V) chain.
A;Reference number: S16024; MUID:91302336; PMID:2071595
A;Accession: S16024
A;Molecule type: mRNA
A;Residues: 1-81,'QL',84-389,'A',391-676,'K',678-1294,'PS',1297,'RS',1300-1553,'R',1555
A;Cross-references: UNIPARC:UPI0000126D1D; GB:D90279; NID:g219509; PIDN:BAA14323.1; PID
A;Note: parts of this sequence were determined by protein sequencing
R;Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.
Exp. Cell Res. 194, 180-185, 1991
A;Title: Insulin binds to type V collagen with retention of mitogenic activity.
A;Reference number: A61142; MUID:91224163; PMID:1709100
A;Accession: A61142
A;Molecule type: protein
A;Residues: 823-824,'X',826-842 <YAO>
A;Cross-references: UNIPARC:UPI0000173BB7
A;Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyanoge
R;Yaoi, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.
Biochim. Biophys. Acta 1035, 139-145, 1990
A;Title: Primary structure of the heparin-binding site of type V collagen.
A;Reference number: S11303; MUID:90366601; PMID:2203476
A;Accession: S11303
A;Molecule type: protein
A;Residues: 823-824,'X',826-848,'I',850-851,'P',853,'PR',856-893,'D',895-932,'X',934-95
A;Cross-references: UNIPARC:UPI0000173BB8
A;Note: the residues designated 'X' are probably glycosylated hydroxylysine; this seque
R;Seyer, J.M.; Kang, A.H.
Arch. Biochem. Biophys. 271, 120-129, 1989
A;Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromide-
A;Reference number: S03978; MUID:89227189; PMID:2496661
A;Accession: S03978
A;Molecule type: protein
A;Residues: 621-640,'G',642-649,'L',651-662,'E',664-667,'Q',669-676,'Q',678-683,'P',685
<SEY>
A;Cross-references: UNIPARC:UPI0000173BB9
A;Note: there are a number of inconsistencies between the sequences in figures 6 and 7;

```
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliaud, M.F.; Boutillon, M.M.; Ber
Eur. J. Biochem. 221, 987-995, 1994
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.
A;Reference number: S43642; MUID:94237164; PMID:8181482
A;Accession: S43642
A;Molecule type: protein
A;Residues: 565-576;756-758,'X',760-763,'X',765-772;1012-1029;1219-1232;1465-1474,'X',14
A;Cross-references: UNIPARC:UPI0000173BBA; UNIPARC:UPI0000173BBB; UNIPARC:UPI0000173BBC;
R;Fessler, L.I.; Brosh, S.; Chapin, S.; Fessler, J.H.
J. Biol. Chem. 261, 5034-5040, 1986
A;Title: Tyrosine sulfation in precursors of collagen V.
A;Reference number: A56977; MUID:86168226; PMID:3082875
A;Contents: annotation; identification of tyrosine sulfate in the amino-terminal propept
R;Lee, S.; Greenspan, D.S.
Biochem. J. 310, 15-22, 1995
A;Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).
A;Reference number: S58665; MUID:95374437; PMID:7646438
A;Accession: S58665
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-36 <LEE>
A;Cross-references: UNIPARC:UPI000007363D; GB:L38808; NID:g1020325; PIDN:AAA79853.1; PID
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
 are 5-hydroxylated and subsequently O-glycosylated.
C;Comment: A long form of the mature protein containing part of the amino-terminal prope
ile the heterotrimers are probably processed to the long form.
C;Genetics:
A;Gene: GDB:COL5A1
A;Cross-references: GDB:131457; OMIM:120215
A;Map position: 9q34.2-9q34.3
C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of t
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
ength, is formed with desmosine cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with cel
A;Note: may play a role in controlling the lateral growth of collagen I fibrils
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F;1-37/Domain: signal sequence #status predicted <SIG>
F;36-261/Domain: PARP-like #status predicted <PARP>
F;38-541/Domain: amino-terminal propeptide #status predicted <PRO>
F;542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>
F;542-558/Region: amino-terminal nonhelical telopeptide
F;559-1572/Region: helical
F;645-647/Region: cell attachment (R-G-D) motif
F;663-665/Region: cell attachment (R-G-D) motif
F;897-929/Region: heparin binding
F;1573-1605/Region: carboxyl-terminal nonhelical telopeptide
F;1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;62-244,183-237/Disulfide bonds: #status predicted
F;159,176,385,1672,1741/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;234,236,240,262,263,273,274,275,277,279,280,338,340,346,347,352,357,416,417,420,421/Bi
F;535/Modified site: allysine (Lys) #status predicted
F;541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F;542/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;570,576,621,639,648,654,657,675,678,690,693,696,705,717,720,726,732,741,750,753,756,76
site: 4-hydroxyproline (Pro) #status experimental
F;627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897/Modified site: 5-hydro
F;627,642,687,774,795,804,807,810,819,825,846,864,882,897,1482/Binding site: carbohydrat
F;708,744/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;1482/Modified site: 5-hydroxylysine (Lys) #status predicted
F;1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted
F;1639,1645,1662,1671/Disulfide bonds: interchain #status predicted
F;1680-1835,1746-1789/Disulfide bonds: #status predicted

  Query Match             13.4%;  Score 280.5;  DB 1;  Length 1838;
  Best Local Similarity   30.1%;  Pred. No. 1.7e-10;
  Matches   83;  Conservative    9;  Mismatches   69;  Indels  115;  Gaps   10;

Qy         13 PAAAPRERGSQGC-GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRE 71
              |   | |:|  |  |  | | | |   |    :|| |

Db        857 PLGPPGEKGKLGVPGLPGYPGRQGPKGS----IGFPGFP---------------------- 891

Qy         72 RGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGH 131
              ||    || |||                          |:| |: |    |
Db        892 -GANGEKGGRGTP---------------------GKPGPRGQRGPTG------------ 916

Qy        132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPP---GPNGPP 188
                          | |    | :      | | | |||          ||||   |||||
Db        917 -------------PRGERGPRGITGKPGPKGNSGGDGP---------AGPPGERGPNGPQ 954

Qy        189 GPPGPPGPQGPPGIPGIPGIPG-------TTVMGPPGPPGP------------------- 222
              || | |||:|||| ||  |:||       |   |  |||||
Db        955 GPTGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGER 1014

Qy        223 --PGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQ 256
                |||  |||| ||  | | | ||:   |    | |:
Db       1015 GHPGPPGPPGEQGLPGLAGKEGTK-GDPGPAGLPGK 1049


RESULT 2
S59856
collagen alpha 1(III) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: S59856; S62120; S16373
R;Toman, P.D.; de Crombrugghe, B.
Gene 147, 161-168, 1994
A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A;Reference number: S59856; MUID:95011609; PMID:7926795
A;Accession: S59856
A;Molecule type: DNA
A;Residues: 1-1464 <TOM>
A;Cross-references: UNIPROT:P08121; UNIPARC:UPI0000177386; EMBL:X52046
R;Toman, D.
submitted to the EMBL Data Library, November 1994
A;Reference number: S62120
A;Accession: S62120
A;Molecule type: DNA
A;Residues: 1-866,'G',868-1464 <TOA>
A;Cross-references: UNIPARC:UPI00000295D6; EMBL:X52046; NID:g575321; PIDN:CAA36279.1; PI
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16373
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1464 <MET>
A;Cross-references: UNIPARC:UPI000016CCAB; EMBL:X57983; NID:g50476; PIDN:CAA41048.1; PID
C;Genetics:
A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 29
58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-154/Domain: propeptide #status predicted <PRO>
F;32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F;155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

  Query Match             13.2%;  Score 277;  DB 2;  Length 1464;
  Best Local Similarity   31.8%;  Pred. No. 2.2e-10;
  Matches   76;  Conservative    6;  Mismatches   75;  Indels   82;  Gaps    8;

Qy         13 PAAAPRERGSQG----CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |   | |||  |     |  ||| : ||  :
Db        792 PRGGPGERGEHGPPGPAGFPGAPGQNGEPGA---------------------------- 822

Qy         69 RRERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSD 126
              : ||||    |  | || :|   | |   |  |  :    | | |      || |
Db        823 KGERGAPGEKGEGGPPGPAGPTGSSGPAGPPGPQGVKGERGSPGP------PGTAG---- 872

Qy        127 SQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNG 186
```

```
                        ||                      |  ||  |       |||||:|
Db        873 -------------FP--------------------GGRGLPGPPGN---NGNPGPPGPSG 896

Qy        187 PPGPPGPPGPQGPPGIPGIPGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
              ||  ||||| |  | || ||| |      | ||  |||| ||||| || | |   | |
Db        897 APGKDGPPGPAGNSGSPGNPGIAGPKGDAGQPGEKGPPGAQGPPGSPGPLGIAGLTGAR 955


RESULT 3
A46053
bullous pemphigoid antigen, BPAG2, type XVII collagen alpha 1-chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 21-Sep-1993 #text_change 09-Jul-2004
C;Accession: A46053
R;Li, K.; Tamai, K.; Tan, E.M.L.; Uitto, J.
J. Biol. Chem. 268, 8825-8834, 1993
A;Title: Cloning of type XVII collagen. Complementary and genomic DNA sequences of mouse
egment, and unusual features in the 5'-end of the gene and the 3'-untranslated region of
A;Reference number: A46053; MUID:93232041; PMID:8473327
A;Accession: A46053
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1433 <LI1>
A;Cross-references: UNIPROT:Q07563; UNIPARC:UPI0000027E06; GB:L08407; NID:g309182; PIDN:
A;Note: sequence extracted from NCBI backbone (NCBIN:129627, NCBIP:129628)

  Query Match             13.1%;  Score 275.5;  DB 2;  Length 1433;
  Best Local Similarity   26.5%;  Pred. No. 2.7e-10;
  Matches   99;  Conservative   32;  Mismatches  114;  Indels  129;  Gaps   16;

Qy         13 PAAAPRERGSQG-------CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELR 65
              || |  ::||:|       | | |  ||:                          :
Db        751 PAGADGQQGSRGEQGLTGMPGTRGPPGPAGDPG-------------------------K 784

Qy         66 SELRRERGAESRLGGSGTPGTSGTLSSLGGL---DPDSPIT--------SHLGQPSPKQQ 114
                |   :| :   |  | ||| |   : | :   :  | ||          :| | |
Db        785 PGLTGPQGPQGLPGSPGRPGTKGEPGAPGRVMTSEGSSTITVPGPPGPPGAMGPPGPPGT 844

Qy        115 PLEPGEAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSK--SNEGAD--GPV 170
              |   | | |    | |          |  :|   :|     :  |:  | :| |  ||
Db        845 PGPAGPAGL--PGQQG----------PRGEPGLAGDSFLSSGSSISEVLSAQGVDLRGP- 891

Qy        171 KNKKKGKKAGPPGPNGPPGP--PGPPGPQGPP--GIPGIPGIPG-------TTVMGPPGP 219
                  ||||| |||||  ||||||:|||  |:|| || ||         |   |||||
Db        892 -------PGPPGPRGPPGPSIPGPPGPRGPPGEGVPGPPGPPGSFLTDSETFFTGPPGP 943

Qy        220 PGPPGPQ---------------------------------GPPGLQGPSGAADKAGTREN 246
              |||||||:                               ||||  || |:   :|
Db        944 PGPPGPKGDQGDPGVPGTPGISGGLSHGASSSTLYMQGPPGPPGPPGPPGSLSSSGQDIQ 1003

Qy        247 QPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQ 306
                  ::|       ::  |||             |      |      : : |  | |||
Db       1004 HYIAEYMQSD----NIRTYLSG---------VQGPPG---PPGPPGPVITITGETFDYSQ 1047

Qy        307 VEVYYINFTDPASY 320
              :   :::   : |
Db       1048 LASQVVSYLRSSGY 1061


RESULT 4
CGBO7S
collagen alpha 1(III) chain - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C;Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
R;Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
A;Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen
A;Reference number: A02862; MUID:80026026; PMID:488906
A;Accession: A02862
A;Molecule type: protein
A;Residues: 1-242 <FIE>
A;Cross-references: UNIPROT:P04258; UNIPARC:UPI0000173B8A
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
A;Title: The covalent structure of calf skin type III collagen. II. The amino acid seque
A;Reference number: A38001; MUID:80026027; PMID:488907
A;Accession: A38001
A;Molecule type: protein
A;Residues: 243-422 <DEW1>
A;Cross-references: UNIPARC:UPI0000173B8B
R;Bentz, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
A;Title: The covalent structure of calf skin type III collagen. III. The amino acid seq
A;Reference number: A38002; MUID:80026028; PMID:488908
A;Accession: A38002
A;Molecule type: protein
A;Residues: 423-571 <BEN>
A;Cross-references: UNIPARC:UPI0000173B8C
R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequ
A;Reference number: A38003; MUID:80026029; PMID:488909
A;Accession: A38003
A;Molecule type: protein
A;Residues: 572-808 <LAN>
A;Cross-references: UNIPARC:UPI0000173B8D
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
A;Title: The covalent structure of calf skin type III collagen. V. The amino acid seque
A;Reference number: A38004; MUID:80026030; PMID:488910
A;Accession: A38004
A;Molecule type: protein
A;Residues: 809-947 <DEW2>
A;Cross-references: UNIPARC:UPI0000173B8E
R;Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequ
A;Reference number: A38005; MUID:80026031; PMID:488911
A;Accession: A38005
A;Molecule type: protein
A;Residues: 948-1049 <ALL>
A;Cross-references: UNIPARC:UPI0000173B8F
A;Experimental source: skin
R;Henkel, W.
Biochem. J. 318, 497-503, 1996
A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
A;Reference number: S71946; MUID:96404897; PMID:8809038
A;Accession: S71946
A;Molecule type: protein
A;Residues: 87-106;1017-1029;1037-1049 <HEN>
A;Cross-references: UNIPARC:UPI0000173B90; UNIPARC:UPI0000173B91; UNIPARC:UPI0000173B92
C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are
C;Comment: The type III collagen molecule is a trimer of identical chains, linked to ea
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
F;1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
F;1-14/Region: amino-terminal nonhelical telopeptide
F;15-1040/Region: helical
F;587-589/Region: cell attachment (R-G-D) motif
F;752-754/Region: cell attachment (R-G-D) motif
F;875-877/Region: cell attachment (R-G-D) motif
F;878-880/Region: cell attachment (R-G-D) motif
F;935-937/Region: cell attachment (R-G-D) motif
F;1041-1049/Region: carboxyl-terminal nonhelical telopeptide
F;95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
F;107,950/Modified site: allysine (Lys) #status predicted
F;107/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;1040,1041/Disulfide bonds: interchain #status predicted

  Query Match             13.1%;  Score 275;  DB 1;  Length 1049;
  Best Local Similarity   31.3%;  Pred. No. 2.1e-10;
  Matches   77;  Conservative    5;  Mismatches   68;  Indels   96;  Gaps    8;

Qy         13 PAAAPRERGSQG----CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |   | ||| ||     |  ||| : ||  :
```

```
Db        637 PRGGPGERGEQGPPGPAGFPGAPGQNGEPGA---------------------------- 667

Qy         69 RRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQ 128

Db        668 KGERGAPGEKGEGGPPGAAGPAGGSGPAGPPGP----QGVKGERGSPGGPGAAG------ 717

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188

Db        718 -----------FP----------------------GGRGP---------PGPPGSNGNP 734

Qy        189 GPP---------GPPGPQGPPGIPGIPGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAA 238

Db        735 GPPGSSGAPGKDGPPGPPGSNGAPGSPGISGPKGDSGPPGERGAPGPQGPPGAPGPLGIA 794

Qy        239 DKAGTR 244

Db        795 GLTGAR 800


RESULT 5
S18803
collagen alpha 1(V) chain - hamster
C;Species: Cricetinae gen. sp. (hamster)
C;Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 16-Dec-1998
C;Accession: S18803
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
J. Biol. Chem. 266, 24727-24733, 1991
A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of e
A;Reference number: S18802; MUID:92105142; PMID:1722213
A;Accession: S18803
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1843 <GRE>
A;Cross-references: UNIPARC:UPI00001773AC
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
F;1620-1842/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

  Query Match            13.0%;  Score 273.5;  DB 2;  Length 1843;
  Best Local Similarity  29.7%;  Pred. No. 4.8e-10;
  Matches   82;  Conservative    9;  Mismatches   70;  Indels  115;  Gaps   10;

Qy         13 PAAAPRERGSQGC-GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRE 71

Db        862 PLGPTGEKGKGLGVPGLPGYPGRQGPKGS----IGFPGFP---------------------- 896

Qy         72 RGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGH 131

Db        897 -GANGEKGGRGTP----------------------GKPGPRGQRGPTG------------ 921

Qy        132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPP---GPNGPP 188

Db        922 -------------PRGERGPRGITGKPGPKGNSGGDGP---------AGPPGERGPNGPQ 959

Qy        189 GPPGPPGPQGPPGIPGIPGIPG-------TTVMGPPGPPGP------------------- 222

Db        960 GPTGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGER 1019

Qy        223 --PGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQ 256

Db       1020 GHPGPPGPPGEQGLPGVAGKEGTK-GDPGPAGLPGK 1054


RESULT 6
A61262
collagen alpha 1(XVII) chain - human (fragment)
N;Alternate names: bullous pemphigoid 180K autoantigen BPAG2; bullous pemphigoid antigen
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C;Accession: I56325; I55345; A61262
R;Giudice, G.J.; Emery, D.J.; Diaz, L.A.
J. Invest. Dermatol. 99, 243-250, 1992
A;Title: Cloning and primary structural analysis of the Bullous pemphigoid autoantigen,
A;Reference number: I56325; MUID:92381323; PMID:1324962
A;Accession: I56325
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1532 <RES>
A;Cross-references: UNIPROT:Q9UMD9; UNIPARC:UPI000006F673; GB:M91669; NID:g179516; PIDN:
R;Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mattei, M.G.; Chu, M.L.; Uitto, J.
J. Biol. Chem. 266, 24064-24069, 1991
A;Title: Genomic organization of collagenous domains and chromosomal assignment of human
A;Reference number: I55345; MUID:92084712; PMID:1748679
A;Accession: I55345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 543-890,'P' <RE2>
A;Cross-references: UNIPARC:UPI000006F72A; GB:M63730; NID:g179520; PIDN:AAA51839.1; PID:
R;Giudice, G.J.; Squiquera, H.L.; Elias, P.M.; Diaz, L.A.
J. Clin. Invest. 87, 734-738, 1991
A;Title: Identification of two collagen domains within the bullous pemphigoid autoantige
A;Reference number: A61262; MUID:91123476; PMID:1846881
A;Accession: A61262
A;Molecule type: mRNA
A;Residues: 543-890,'P' <GIU>
A;Cross-references: UNIPARC:UPI000006F72A
C;Genetics:
A;Gene: GDB:COL17A1; BPAG2; BP180
A;Cross-references: GDB:131396; OMIM:113811
A;Map position: 10q24.3-10q24.3

  Query Match            13.0%;  Score 273;  DB 2;  Length 1532;
  Best Local Similarity  31.2%;  Pred. No. 4.3e-10;
  Matches   85;  Conservative   13;  Mismatches   64;  Indels  110;  Gaps   13;

Qy         20 RGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRERGAE 75

Db        789 RGEQGLTGMPGIRGPPGPSGDP-------GKPGLT----------------------GPQ 819

Qy         76 SRLGGSGTPGTSGTLSSLGGLDPDSPITSH----------------LGQPSPKQQPLEPG 119

Db        820 GPQGLPGTPGRPGIKGEPGA--PGKIVTSEGSSMLTVPGPPGPPGAMGPPGPPGAPGPAG 877

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKA 179

Db        878 PAGL-----PGHQEVL----------------------------NLQGP---------P 894

Qy        180 GPPGPNGPPGP--PGPPGPQGPP--GIPGIPGIPG-------TTVMGPPGPPGPPGP--- 225

Db        895 GPPGPRGPPGPSIPGPPGPRGPPGEGLPGPPGPPGSFLSNSETFLFGPPGPPGPPGPKGD 954

Qy        226 QGPPGLQGPSGAADKAG--TRENQPAVVHLQG 255

Db        955 QGPPGPRGHQGEQGLPGFSTSGSSSFGLNLQG 986


RESULT 7
JC7595
scavenger receptor with C-type lectin type I - human
C;Species: Homo sapiens (man)
C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C;Accession: JC7595
R;Nakamura, K.; Funakoshi, H.; Miyamoto, K.; Tokunaga, F.; Nakamura, T.
Biochem. Biophys. Res. Commun. 280, 1028-1035, 2001
A;Title: Molecular cloning and functional characterization of a human scavenger receptor
A;Reference number: JC7595; MUID:21092718; PMID:11162630
A;Contents: Placenta
A;Accession: JC7595
A;Molecule type: mRNA
A;Residues: 1-742 <NAK>
A;Cross-references: UNIPROT:Q9BYH7; UNIPARC:UPI000006F955; DDBJ:AB038518
C;Comment: This receptor, a member of the scavenger receptor family, belonging to the ty
 important role in host defense. It forms a timer and plays a role in recognizing infect
C;Genetics:
A;Gene: srcl-I
A;Map position: 18p11.32
C;Keywords: coiled coil; glycoprotein; transmembrane protein
```

```
F;1-39/Domain: cytosolic (amino-terminus) #status predicted <CYT>
F;16-19/Region: internalization signal YKRF
F;40-56/Domain: transmembrane #status predicted <TMM>
F;57-112/Domain: extracellular #status predicted <EXT>
F;113-335/Domain: coiled coil #status predicted <COC>
F;369-384/Region: serine/threonine-rich #status predicted
F;443-589/Domain: collagen-like #status predicted <COL>
F;607-732/Domain: C-type lectin/carbohydrate recognition domain #status predicted <CRD>

  Query Match            12.8%;  Score 268.5;  DB 2;  Length 742;
  Best Local Similarity   37.2%;  Pred. No. 4e-10;
  Matches   67;  Conservative   13;  Mismatches   63;  Indels   37;  Gaps    7;

Qy         82 GTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNFFFP 141
              | |||  |      |   |  | | : ||   | :|  ||| |
Db        443 GPPGPRGPRGDRGSQGPPGP-TGNKGQKGEKGEPGPPPGPAG------------------- 482

Qy        142 DEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGP------PG 195
              :  |          |  | ||   |: ||  ::    | || ||:| ||||||      ||
Db        483 ERGPIGPAGPPGERGGKGSK---GSQGPKGSRGSPGKPGPQGPSGDPGPPGPPGKEGLPG 539

Qy        196 PQGPPGIPGI------PGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQP 248
              ||||||  |:      ||:||   : | || || |||:|||| |||||       :|:|
Db        540 PQGPPGPFQGLQGTVGEPGVPGPRGLPGLPGVPGMPGPKGPPGPPGPSGAVVPLAL-QNEP 598


RESULT 8
T20497
hypothetical protein F02D10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T20497
R;Swinburne, J.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19283
A;Accession: T20497
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-316 <WIL>
A;Cross-references: UNIPROT:Q19111; UNIPARC:UPI0000080924; EMBL:Z67990; PIDN:CAA91932.1;
A;Experimental source: clone F02D10
C;Genetics:
A;Gene: CESP:F02D10.1
A;Map position: X
A;Introns: 56/3

  Query Match            12.8%;  Score 267.5;  DB 2;  Length 316;
  Best Local Similarity   32.8%;  Pred. No. 1.9e-10;
  Matches   78;  Conservative    3;  Mismatches   82;  Indels   75;  Gaps    8;

Qy         21 GSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRERGAESRLGG 80
              |: | | ||    || |  |                | ||           |   | |
Db         94 GAGGGGGGGGGEAAGGGGC----------------TGCCN-------PGPPGPGGRPGK 130

Qy         81 SGTPGTSGT-----LSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMAL 135
               |||| |       |  |   | | |     || |   |  ||
Db        131 PGTPGKPGAPGNPGASGKGAAAPCEPSTPPPCQPCPAGPPGPPG---------------- 174

Qy        136 LNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP--PGPNGPPGPPGP 193
                 ||  |                 |  ||      |  |||  ||| ||||| ||
Db        175 -----PDGTP---------------GEPGGPGPA-----GSPAGPSGPGPAGPPGPAGP 208

Qy        194 PGPQGPPGIPGIPGIPGTTVMG---PPGPPGPPGPQGPPGLQGPSGAADKAGTRENQP 248
              ||  | || || ||  | :  |    ||| ||||| || |  | ||:    |      |
Db        209 PGNDGQPGQPGGPGQDGASSAGGEAGPGPAGPPGPAGPAGPDGQSGSGSAGGPGPKGP 266


RESULT 9
I50694
collagen alpha 1(III) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I50694
R;Nah, H.D.; Niu, Z.; Adams, S.L.
J. Biol. Chem. 269, 16443-16448, 1994
A;Title: An alternative transcript of the chick type III collagen gene that does not enc
A;Reference number: A54041; MUID:94266842; PMID:8206952
A;Accession: I50694
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-886 <NAH>
A;Cross-references: UNIPROT:P12105; UNIPARC:UPI000017125A; EMBL:U07973; NID:g520454; PID
C;Genetics:
A;Gene: COL3A1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;30-90/Domain: von Willebrand factor type C repeat homology <VWC>

  Query Match            12.7%;  Score 267;  DB 2;  Length 886;
  Best Local Similarity   32.5%;  Pred. No. 6e-10;
  Matches   92;  Conservative   16;  Mismatches   95;  Indels   80;  Gaps   14;

Qy          4 PEVERREL-LPAA--APRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLL 56
              |: || |   | |   | | | :|     |  | |   ||  |     || ||  :
Db        444 PKGERGENGTPGARGPPGEEGKRGANGEPGQNGVPGTPGERGS----PGFRGLPGS---- 495

Qy         57 TLCCYLELRSELRRERGAESRLGGSGTPGTSGT-----------LSSLGGLD--PDSPIT 103
                  : |  |:|      |  | | || ||             |  : ||    | || :
Db        496 ---------NGLPGEKGPAGERGSPGPPGPSGPAGDRGQDGGPGLPGMRGLPGIPGSPGS 546

Qy        104 SHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSN 163
                |:| |     ||| :     :    |  :: | |                    | |
Db        547 D--GKPGPPGNQGEPGRSGPPGPAGPRGQPGVMGFPGP------------------KGN 585

Qy        164 EGADGPVKNKKKGKKAGPPGPNGP---------PGPPGP---------PGPQGPPGIPGI 205
              ||| |  || ::|   ||||  ||          ||||||             ||| | ||: |:
Db        586 EGAPG--KNGERG-PGGPPGTPGPAGKNGDVGLPGPPGPAGPAGDRGEPGPSGSPGLQGL 642

Qy        206 PGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQ 247
              || ||     | || ||| |  | ||  || |       | | |
Db        643 PGGPGPAGENGKPGEPGPKGDIGGPGFPGPKGENGIPGERGPQ 685


RESULT 10
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N;Contains: endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: A56101; A58371; S72450; S65595
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tiss
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: A56101
A;Molecule type: mRNA
A;Residues: 1-103 <REH1>
A;Cross-references: UNIPROT:P39061; UNIPARC:UPI000016CAD0; GB:U11636; NID:g618427; PIDN
R;Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous
A;Reference number: A58371; MUID:94240112; PMID:8183894
A;Accession: A58371
A;Molecule type: mRNA
A;Residues: 1-928 <REH2>
A;Cross-references: UNIPARC:UPI000016CCAD; GB:L16898; NID:g404754; PIDN:AAA37434.1; PID
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 28-687,'L',689-734,'F',736-751,'R',753-1315 <OHW>
A;Cross-references: UNIPARC:UPI000016CCB8; EMBL:L22545; NID:g348968; PIDN:AAA19787.1; P
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xa
```

```
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OHS>
A;Cross-references: UNIPARC:UPI000017A14D; EMBL:L22545
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of ur
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal homologous
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F;327-353/Domain: collagenous #status predicted <CO1>
F;364-437/Domain: collagenous #status predicted <CO2>
F;462-583/Domain: collagenous #status predicted <CO3>
F;607-689/Domain: collagenous #status predicted <CO4>
F;704-745/Domain: collagenous #status predicted <CO5>
F;759-831/Domain: collagenous #status predicted <CO6>
F;842-874/Domain: collagenous #status predicted <CO7>
F;887-910/Domain: collagenous #status predicted <CO8>
F;892-894/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <CO9>
F;983-1000/Domain: collagenous #status predicted <CO10>
F;1132-1315/Product: endostatin #status predicted <EST>
F;1139-1315/Region: multiplexin collagen carboxyl-terminal homologous
F;126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172-228/Disulfide bonds: #status predicted
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             12.7%;  Score 267;  DB 2;  Length 1315;
  Best Local Similarity   30.2%;  Pred. No. 9e-10;
  Matches   71;  Conservative   17;  Mismatches   55;  Indels   92;  Gaps    9;

Qy          72 RGAESRLGGSGTPGTSGTLSSLGGL--DPDS-PITSHLGQPSPK------------QQPL 116
               :| :   |  | || :|      | :   |:| |:    | | |:             :|
Db         329 KGQKGEPGAQGPPGPAGPQGPAGPVVQSPNSQPVPGAQGPPGPQGPPGKDGTPGRDGEPG 388

Qy         117 EPGEAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKG 176
               :|||     |:  | |       ||                    |  || | :|
Db         389 DPGEDGRPGDT--GPQ------GFP--------------------GTPGDVGP-KGEKGD 419

Qy         177 KKAGPPGPNGPPGPPGP-----------------------------PGPQGPPGIPGIP 206
                  || || ||||||||                            ||| ||||:||:|
Db         420 PGIGPRGPPGPPGPPGPSFRQDKLTFIDMEGSGFSGDIESLRGPRGFPGPPGPPGVPGLP 479

Qy         207 GIPGTTVM-----------------GPPGPPGPPGPQGPPGLQGPSGAADKAGT 243
               | ||   :                 |||| |||||| |||| :|| | | : |:
Db         480 GEPGRFGINGSYAPGPAGLPGVPGKEGPPGPPGPPGPPGPPGKEGPPGVAGQKGS 534


RESULT 11
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
N;Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: B56101
A;Molecule type: mRNA
A;Residues: 1-562 <REH1>
A;Cross-references: UNIPROT:Q61434; UNIPARC:UPI000016CAD1; GB:U11637; NID:g618429; PIDN:
A;Experimental source: splice form clone PE17.24
A;Accession: C56101
A;Molecule type: mRNA
A;Residues: 1-239,487-562 <REH2>
A;Cross-references: UNIPARC:UPI000017A14C; GB:U11637; NID:g618429
A;Experimental source: splice form clones PE8.1, PE19, PE15.2
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 487-1146,'L',1148-1193,'F',1195-1210,'R',1212-1512,'L',1514-1522,'F',1524-16
A;Cross-references: UNIPARC:UPI000016CCB8; EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PI
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 487-1512,'L',1514-1522,'F',1524-1683,'V',1685-1774 <OH2>
A;Cross-references: UNIPARC:UPI000017A14D; EMBL:L22545
R;Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A;Title: Identification of a novel collagen chain represented by extensive interruptions
A;Reference number: PN0675; MUID:94059075; PMID:8240330
A;Accession: PN0675
A;Molecule type: mRNA
A;Residues: 635-1774 <ABE>
A;Cross-references: UNIPARC:UPI00000E7D60
R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str
collagen chain.
A;Reference number: A54072; MUID:94245707; PMID:8188673
A;Accession: A54072
A;Molecule type: DNA; mRNA
A;Residues: 1293-1403,'R',1405-1774 <REH3>
A;Cross-references: UNIPARC:UPI000016CAAB; GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bi
Cell 88, 277-285, 1997
A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A;Reference number: A58816; MUID:97160848; PMID:9008168
A;Accession: A58816
A;Molecule type: protein
A;Residues: 1591-1610 <ORE>
A;Cross-references: UNIPARC:UPI00000499D2
A;Experimental source: hemangioendothelium cells
A;Note: inhibits endothelial cell proliferation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of ur
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
A;Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/
A;Note: the list of introns is incomplete
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pred
F;1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form #s
F;1-24/Domain: signal sequence #status predicted <SIG>
F;786-812/Domain: collagenous #status predicted <CO01>
F;823-896/Domain: collagenous #status predicted <CO02>
F;921-1042/Domain: collagenous #status predicted <CO03>
F;1066-1148/Domain: collagenous #status predicted <CO04>
F;1163-1204/Domain: collagenous #status predicted <CO05>
F;1218-1290/Domain: collagenous #status predicted <CO06>
F;1301-1333/Domain: collagenous #status predicted <CO07>
F;1346-1369/Domain: collagenous #status predicted <CO08>
F;1351-1353/Region: cell attachment (R-G-D) motif
F;1377-1428/Domain: collagenous #status predicted <CO09>
```

F;1442-1459/Domain: collagenous #status predicted <CO10>
F;1591-1774/Product: endostatin #status predicted <EST>
F;1598-1774/Region: multiplexin collagen carboxyl-terminal homologous
F;354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 12.7%; Score 267; DB 2; Length 1774;
Best Local Similarity 30.2%; Pred. No. 1.2e-09;
Matches 71; Conservative 17; Mismatches 55; Indels 92; Gaps 9;

```
Qy          72 RGAESRLGGSGTPGTSGTLSSLGGL--DPDS-PITSHLGQPSPK-----------QQPL 116
               :| :   |  | || :|     | :   |:| |:    | | |:           :|
Db         788 KGQKGEPGAQGPPGPAGPQGPAGPVVQSPNSQPVPGAQGPPGPQGPPGKDGTPGRDGEPG 847

Qy         117 EPGEAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKG 176
               :|||     |:  | |        ||                          |  || | :|
Db         848 DPGEDGRPGDT--GPQ------GFP-------------------GTPGDVGP-KGEKGD 878

Qy         177 KKAGPPGPNGPPGPPGP----------------------------PGPQGPPGIPGIP 206
                 || || ||||||||                            ||| ||||:||:|
Db         879 PGIGPRGPPGPPGPPGPSFRQDKLTFIDMEGSGFSGDIESLRGPRGFPGPPGPPGVPGLP 938

Qy         207 GIPGTTVM-----------------GPPGPPGPPGPQGPPGLQGPSGAADKAGT 243
               | ||   :                 |||| |||||| |||| :|| | | : |:
Db         939 GEPGRFGINGSYAPGPAGLPGVPGKEGPPGFPGPPGPPGPPGKEGPPGVAGQKGS 993
```

RESULT 12
CGBO1S
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048; A02853
R;Rauterberg, J.; Timpl, R.; Furthmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A;Title: Structural characterization of N-terminal antigenic determinants in calf and hu
A;Reference number: A91193; MUID:72255334; PMID:4115172
A;Accession: A91193
A;Molecule type: protein
A;Residues: 1-19 <RAU>
A;Cross-references: UNIPROT:P02453; UNIPARC:UPI0000173B51
A;Experimental source: skin
A;Note: the epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain, is conve
R;Pietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A;Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide
A;Reference number: A91229; MUID:76022320; PMID:1164916
A;Accession: A91229
A;Molecule type: protein
A;Residues: 20-145 <FIE>
A;Cross-references: UNIPARC:UPI0000173B52
A;Experimental source: skin
A;Note: Lys-103 is hydroxylated and binds glucosylgalactose
R;Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A;Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3 from calf
A;Reference number: A91387; MUID:73049499; PMID:4673951
A;Accession: A91387
A;Molecule type: protein
A;Residues: 146-294 <FI2>
A;Cross-references: UNIPARC:UPI0000173B53
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 396-400, 1973
A;Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fr
A;Reference number: A91211; MUID:74086118; PMID:4359390
A;Accession: A91211
A;Molecule type: protein
A;Residues: 295-562 <FI3>
A;Cross-references: UNIPARC:UPI0000173B54
A;Experimental source: skin
R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972
A;Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
A;Reference number: A91201; MUID:73042276; PMID:4343808
A;Accession: A91201
A;Molecule type: protein
A;Residues: 563-675 <WEN>
A;Cross-references: UNIPARC:UPI0000173B55
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A;Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C
A;Reference number: A91200; MUID:73042275; PMID:4343807
A;Accession: A91200
A;Molecule type: protein
A;Residues: 676-758 <FI4>
A;Cross-references: UNIPARC:UPI0000173B56
A;Experimental source: skin
A;Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in positi
R;Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
FEBS Lett. 21, 75-79, 1972
A;Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of
A;Reference number: A43048
A;Accession: A43048
A;Molecule type: protein
A;Residues: 759-779 <RA2>
A;Cross-references: UNIPARC:UPI0000173B57
A;Experimental source: skin
C;Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydro
C;Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are
C;Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine skin
9, 149, 268, and 217 residues.
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 12.7%; Score 266; DB 1; Length 779;
Best Local Similarity 30.0%; Pred. No. 6.1e-10;
Matches 78; Conservative 16; Mismatches 80; Indels 86; Gaps 9;

```
Qy          13 PAAAPRERGSQG----CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCYLEL 64
               || || ::|  |      |  |||   ||    | :     || |
Db         339 PAGAPGDKGEAGPSGPAGTRGAPGDRGEPGPPGPA-----GFAG--------------- 377

Qy          65 RSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALH 124
                  ||: : |  | || :|     |   |  |     | | |       ||
Db         378 ------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGP---- 423

Qy         125 SDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGP 184
                                                  | :: : |  |          : |||||
Db         424 --------------------------------KGARGSAGPPGATGFPGAAGRVGPPGP 450

Qy         185 NGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAA 238
               :|  ||||||||      :|| |  |  | ||   :||||||| | :| ||  ||:||
Db         451 SGNAGPPGPPGPAGKEGSKGPRGETGPAGRPGE--VGPPGPPGPAGEKGAPGADGPAGAP 508

Qy         239 DKAGTR--ENQPAVVHLQGQ 256
                  | :    |  || | ||
Db         509 GTPGPQGIAGQRGVVGLPGQ 528
```

RESULT 13
CGHU1S
collagen alpha 1(I) chain precursor - human
N;Alternate names: procollagen alpha 1(I) chain
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 31-Dec-2004
C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1
5269; A29439; I53466; A02852; I37247
R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorious, P.J
Gene 67, 105-115, 1988
A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five
A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-369,'L',371-589 <DAL>
A;Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP
R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock
Biochem. J. 253, 919-922, 1988
A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human
A;Reference number: S01143; MUID:89025644; PMID:3178743
A;Accession: S01143
A;Molecule type: mRNA
A;Residues: 1-472 <TRO>
A;Cross-references: UNIPARC:UPI000016A6F9; EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID
A;Note: submitted to the EMBL/GenBank/DDBJ databases by Prockop, D.J., 13-JUN-1988
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;
Nature 310, 337-340, 1984
A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A;Reference number: A93335; MUID:84270697; PMID:6462220
A;Accession: A93335
A;Molecule type: DNA
A;Residues: 1-58,'Q',60-181 <CHU>
A;Cross-references: UNIPARC:UPI0000173B3C; EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID
R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 262, 15151-15157, 1987
A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh
A;Reference number: I55254; MUID:88033098; PMID:2822714
A;Accession: I55254
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-45 <ROS>
A;Cross-references: UNIPARC:UPI000016A6B2; GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:
R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A;Title: Regulatory elements in the first intron contribute to transcriptional control o
A;Reference number: A39943; MUID:88097389; PMID:3480516
A;Accession: A39943
A;Molecule type: DNA
A;Residues: 1-34 <BOR>
A;Cross-references: UNIPARC:UPI000016A6B1; GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A;Reference number: I55237; MUID:85130970; PMID:2857713
A;Accession: I55237
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: UNIPARC:UPI000016A6B1; GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A;Reference number: A35233; MUID:90202908; PMID:2318855
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WIR>
A;Cross-references: UNIPARC:UPI0000173B3D
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
A;Reference number: S09400; MUID:89356643; PMID:2767050
A;Accession: S09400
A;Molecule type: mRNA
A;Residues: 156-183 <WEI>
A;Cross-references: UNIPARC:UPI0000173B3E
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A;Reference number: A90567; MUID:71038625; PMID:5529814
A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 162-198,'Z',200-201,'Z',203-206,'Z',208-209,'Z',211-228,'B',230,'BB',233,'Z'
A;Cross-references: UNIPARC:UPI0000173B3F
A;Experimental source: skin
A;Note: evidence for 170-allysine
R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, P
Eur. J. Biochem. 192, 153-159, 1990
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A;Reference number: S11372; MUID:90382436; PMID:2169412
A;Accession: S11372
A;Molecule type: protein
A;Residues: 175-187;274-287,'P',289 <BAE>
A;Cross-references: UNIPARC:UPI0000173B40; UNIPARC:UPI0000173B41
A;Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
ooperative melting of intact type I collagen.
A;Reference number: I55342; MUID:92042092; PMID:1718984
A;Accession: I55342
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 258-268;1347-1357 <DEA>
A;Cross-references: UNIPARC:UPI0000173B42; UNIPARC:UPI0000173B43; GB:S67495; NID:g239007
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A;Reference number: A92069; MUID:71001508; PMID:4319110
A;Accession: A92069
A;Molecule type: protein
A;Residues: 263-268 <MOR>
A;Cross-references: UNIPARC:UPI000014DF11
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A;Title: Segmental amplification of the entire helical and telopeptide regions of the cD
A;Reference number: S15989; MUID:90326017; PMID:2374517
A;Accession: S15989
A;Molecule type: mRNA
A;Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>
A;Cross-references: UNIPARC:UPI0000173B44; UNIPARC:UPI0000173B45; UNIPARC:UPI0000173B46;
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A;Reference number: I52905; MUID:93339042; PMID:8339541
A;Accession: I52905
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 342-352,'C',354-359 <WI2>
A;Cross-references: UNIPARC:UPI0000070EE6; GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:
A;Note: mutant sequence from patient with osteogenesis imperfecta
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
A;Reference number: A90476; MUID:84080385; PMID:6689127
A;Accession: A90476
A;Molecule type: mRNA
A;Residues: 425-1250,'X',1252-1328,'S',1330-1390,'X',1392-1464 <BER>
A;Cross-references: UNIPARC:UPI0000173B4A; GB:K01228; NID:g180391; PIDN:AAA51995.1; PID
A;Note: sequence partially completed for missing nucleotides by A29439
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multiexon deletion in an osteogenesis imperfecta variant with increased type II
A;Reference number: A22161; MUID:85104934; PMID:2981843
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594,'R',596-607 <CH3>
A;Cross-references: UNIPARC:UPI000011F796; GB:K03178; GB:K03179; NID:g179612; NID:g17961
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained

A;Reference number: A35336; MUID:90252792; PMID:2339700
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720,'E',722-737,'E',739-745 <WAL>
A;Cross-references: UNIPARC:UPI0000173B4B
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
A;Reference number: I54365; MUID:95187161; PMID:7881420
A;Accession: I54365
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 746-766,'S',768-781 <FOR>
A;Cross-references: UNIPARC:UPI000016A6FA; GB:L47667; NID:g1009093; PIDN:AAB59576.1; PID
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A;Reference number: A47426; MUID:93352646; PMID:8349697
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1336,1339-1387,'R',1389-1464 <CHE>
A;Cross-references: UNIPARC:UPI0000073A2A; GB:S64596; NID:g407589; PIDN:AAB27856.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIP:136445)
A;Note: does not represent an experimentally determined sequence but three different mut
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Cross-references: UNIPARC:UPI0000173B4C
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1464 <CH5>
A;Cross-references: UNIPARC:UPI0000173B4D
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336,1339-1464 <CH6>
A;Cross-references: UNIPARC:UPI0000173B4E
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387,'R',1389-1464 <CH7>
A;Cross-references: UNIPARC:UPI0000173B4F
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nic
J. Biol. Chem. 263, 14605-14607, 1988

Query Match 12.7%; Score 265.5; DB 1; Length 1464;
Best Local Similarity 29.7%; Pred. No. 1.3e-09;
Matches 78; Conservative 16; Mismatches 80; Indels 89; Gaps 9;

```
Qy          13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
               || || ::|  |         |  |||   ||    | :    || |
Db         774 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------ 815

Qy          62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
                  ||: : |  | || :|     |  |  |     | | |       ||
Db         816 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 861

Qy         122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                  | :: : |  |        : ||
Db         862 ------------------------------------KGARGSAGPPGATGFPGAAGRVGP 885

Qy         182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
               |||:|  ||||||||      :|| |  |  | ||   :||||||| | :| ||  ||:
Db         886 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 943

Qy         236 GAADKAGTR--ENQPAVVHLQGQ 256
               ||    | :    |  || | ||
Db         944 GAPGTPGPQGIAGQRGVVGLPGQ 966
```

RESULT 14
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C;Species: Ascaris suum (pig roundworm)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
C;Accession: S16366
R;Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the partia
A;Reference number: S16366; MUID:91340768; PMID:1714907
A;Accession: S16366
A;Molecule type: mRNA
A;Residues: 1-1763 <JBI>
A;Cross-references: UNIPROT:P27393; UNIPARC:UPI0000126D40; GB:M67507; NID:g159648; PIDN:
C;Genetics:
A;Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F;27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F;43-1529/Domain: collagenous #status predicted <COL>
F;197-199/Region: cell attachment (R-G-D) motif
F;1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1530-1638/Domain: repeat NC1 #status predicted <NC11>
F;1639-1763/Domain: repeat NC1 #status predicted <NC12>
F;31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 12.6%; Score 265; DB 2; Length 1763;
Best Local Similarity 34.3%; Pred. No. 1.6e-09;
Matches 85; Conservative 15; Mismatches 82; Indels 66; Gaps 10;

```
Qy          13 PAAAPRERGSQG-------CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELR 65
               |  ||  :| :|         | | |  ||:       |  |:     |
Db         256 PEGAPGMKGEKGDPGEAGPRGFPGTPGVAGQP-------GLPGMKGEKGL---------- 298

Qy          66 SELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHS 125
               |      || | | |  | || |      |||          | |      | : |||
Db         299 SGPAGPRGKEGRPGLPGPPGFKGD----RGLD---------GLPGVPGLPGQKGEAGF-- 343

Qy         126 DSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPN 185
                 :|| : |       |   |   | |             :|  ||     :  : |||| :
Db         344 PGRDGAKGAR----GPPGPPGGGEFS-----------DGPPGPPGLPGREGQPGPPGAD 387

Qy         186 GPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGP---------PGPPGPQGPPGLQGPSG 236
               | |||||| |||| || ||:||:||  : | |||          || || |||||| || |
Db         388 GYPGPPGPQGPQGLPGGPGLPGLPG--LEGLPGPKGEKGDSGIPGAPGVQGPPGLAGPPG 445

Qy         237 AADKAGTR 244
               |  : | |
Db         446 AKGEPGPR 453
```

RESULT 15
CGHU1B
collagen alpha 4(IV) chain precursor - human
N;Alternate names: procollagen alpha 4(IV) chain
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004
C;Accession: A55360; S36854; S28777
R;Leinonen, A.; Mariyama, M.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.
J. Biol. Chem. 269, 26172-26177, 1994
A;Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. Comp
A;Reference number: A55360; MUID:95014445; PMID:7523402
A;Accession: A55360
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1690 <LEI>
A;Cross-references: UNIPROT:P53420; UNIPARC:UPI0000000769; GB:X81053; NID:g574805; PIDN:
R;Sugimoto, M.; Oohashi, T.; Yoshioka, H.; Matsuo, N.; Ninomiya, Y.

```
FEBS Lett. 330, 122-128, 1993
A;Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen cha
A;Reference number: S36854; MUID:93374047; PMID:8365481
A;Accession: S36854
A;Molecule type: DNA; mRNA
A;Residues: 1219-1658,'FE',1661-1690 <SUG>
A;Cross-references: UNIPARC:UPI000016A432; DDBJ:D17391; NID:g440365; PIDN:BAA04214.1; PI
A;Experimental source: whole eye
R;Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
A;Reference number: S28777; MUID:93054733; PMID:1429714
A;Accession: S28777
A;Molecule type: DNA
A;Residues: 1407-1424,'G',1426-1430,'A',1432-1439,'L',1441-1507 <KAM>
A;Cross-references: UNIPARC:UPI0000173BFC; GB:L01475; GB:L01476
A;Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with the
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A4
A;Cross-references: GDB:132673; OMIM:120131
A;Map position: 2q35-2q37
A;Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3(
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-li
C;Function:
A;Description: minor structural component of extracellular basement membrane in kidney g
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F;1-38/Domain: signal sequence #status predicted <SIG>
F;39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
F;39-61/Domain: amino-terminal nonhelical, NH1 <NH1>
F;62-1466/Region: interrupted helical
F;94-96/Region: cell attachment (R-G-D) motif
F;145-147/Region: cell attachment (R-G-D) motif
F;189-191/Region: cell attachment (R-G-D) motif
F;310-312/Region: cell attachment (R-G-D) motif
F;724-726/Region: cell attachment (R-G-D) motif
F;785-787/Region: cell attachment (R-G-D) motif
F;989-991/Region: cell attachment (R-G-D) motif
F;1212-1214/Region: cell attachment (R-G-D) motif
F;1467-1690/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1579-1686/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;47,52,55,57,266,400,460,492,494,668,790,828,1095,1131,1294,1317,1375,1407/Disulfide bo
F;142,669/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1480-1566,1513-1569/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
F;1525-1531,1634-1641/Disulfide bonds: #status predicted
F;1588-1683,1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted

  Query Match            12.6%;  Score 264;  DB 1;  Length 1690;
  Best Local Similarity  27.4%;  Pred. No. 1.8e-09;
  Matches   96;  Conservative   18;  Mismatches   96;  Indels  140;  Gaps   15;

Qy          2 GYPEVERRELLPAAAPRERGSQG-CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              | |   | |       | |:|  |  |  | |   |        |:|| ||
Db       1008 GPPGFHRGE------PGEKGQPGPPGPPGPPGSTGLRG----FIGFPGLP---------- 1047

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPG- 119
                | :   |  | || ||   : |          || | | || :|  ||
Db       1048 ------------GDQGEPGSPGPPGPFSGIDGARGPKGNKGDPASHFGPPGPKGEPGSPGC 1095

Qy        120 ---------------EAALHSDSQDG------------HQMALLNFFFPDEK------- 144
                            :   |  : |            | |  |   | |
Db       1096 PGHFGASGEQGLPGIQGPRGSPGRPGPPGSSGPPGCPGDHGMPGLR-GQPGEMGDPGPRG 1154

Qy        145 ---------------PYSEEESRRVRRNKRSKSNEGAD--------GPV-----KNKK-- 174
                            |         :   |  | :||         |||     | ::
Db       1155 LQGDPGIPGPPGIKGPSGSPGLNGLHGLKGQKGTKGASGLHDVGPPGPVGIPGLKGERGD 1214

Qy        175 -----------KGKKA--GPPGPNGPPGP------------PGPPGPQGPP------GI 202
                        :|||   |||| :|||||             |||| ||||      |
Db       1215 PGSPGISPPGPRGKKGPPGPPGSSGPPGPAGATGRAPKDIPDPGPPGDQGPPGPDGPRGA 1274

Qy        203 PGIPGIPGTTVM--------GPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
              || ||:||:  :        | |||||||||| |||| :|  |   | | :
Db       1275 PGPPGLPGSVDLLRGEPGDCGLPGPPGPPGPPGPPGYKGFPGCDGKDGQK 1324


Search completed: January 13, 2006, 16:15:22
Job time : 17 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2006, 16:07:11 ; Search time 73 Seconds
(without alignments)
3778.926 Million cell updates/sec

Title: US-09-729-658C-2
Perfect score: 2098
Sequence: 1 MGYPEVERRELLPAAAPRER.....NMSKHTTFFGAIRLGEAPAS 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2098	100.0	391	1	EDA_HUMAN	Q92838 homo sapien
2	1972	94.0	391	1	EDA_BOVIN	Q9beg5 bos taurus
3	1965	93.7	391	1	EDA_MOUSE	O54693 mus musculu
4	1890.5	90.1	386	2	Q58P77_CANFA	Q58p77 canis famil
5	1441.5	68.7	356	2	Q5EFZ5_CHICK	Q5efz5 gallus gall
6	955.5	45.5	367	2	Q50D54_GASAC	Q50d54 gasterosteu
7	887	42.3	178	2	Q5JS00_HUMAN	Q5js00 homo sapien
8	809	38.6	207	2	Q4SJR1_TETNG	Q4sjr1 tetraodon n
9	286	13.6	742	2	Q8VIF6_MOUSE	Q8vif6 mus musculu
10	284	13.5	998	2	Q8CFM4_MOUSE	Q8cfm4 mus musculu
11	282	13.4	742	2	Q8K4Q8_MOUSE	Q8k4q8 mus musculu
12	282	13.4	742	2	Q8C979_MOUSE	Q8c979 mus musculu
13	280.5	13.4	1792	2	Q59EE7_HUMAN	Q59ee7 homo sapien
14	280.5	13.4	1838	1	CO5A1_HUMAN	P20908 homo sapien
15	280.5	13.4	1838	2	Q15094_HUMAN	Q15094 homo sapien
16	280.5	13.4	1838	2	Q5SUX4_HUMAN	Q5sux4 homo sapien
17	280.5	13.4	1840	2	Q59IP3_PIG	Q59ip3 sus scrofa
18	280	13.3	284	2	Q25581_9BILA	Q25581 teladorsagi
19	278	13.3	742	2	Q4V885_RAT	Q4v885 rattus norv
20	277.5	13.2	1838	2	O88207_MOUSE	O88207 mus musculu
21	277.5	13.2	1840	2	Q9JI03_RAT	Q9ji03 rattus norv
22	276	13.2	622	2	Q9BY85_HUMAN	Q9by85 homo sapien
23	276	13.2	1315	2	Q8QHL9_XENLA	Q8qhl9 xenopus lae
24	275.5	13.1	1470	1	COHA1_MOUSE	Q07563 mus musculu
25	275	13.1	1049	1	CO3A1_BOVIN	P04258 bos taurus
26	275	13.1	1431	1	COHA1_MESAU	Q9jmh4 mesocricetu
27	273.5	13.0	1840	2	Q60467_CRILO	Q60467 cricetulus
28	273	13.0	742	2	Q5KU26_HUMAN	Q5ku26 homo sapien
29	273	13.0	742	2	Q8WZA4_HUMAN	Q8wza4 homo sapien
30	273	13.0	1497	1	COHA1_HUMAN	Q9umd9 homo sapien
31	272	13.0	284	2	Q25582_9BILA	Q25582 teladorsagi
32	272	13.0	1415	2	Q5JV36_HUMAN	Q5jv36 homo sapien
33	272	13.0	1802	2	Q17163_BRUMA	Q17163 brugia mala
34	271.5	12.9	1344	2	O93419_CHICK	O93419 gallus gall
35	271	12.9	326	2	Q5C1A3_SCHJA	Q5c1a3 schistosoma
36	270	12.9	1222	2	Q8K173_MOUSE	Q8k173 mus musculu
37	270	12.9	1464	1	CO3A1_MOUSE	P08121 mus musculu
38	270	12.9	1464	2	Q8BLW4_MOUSE	Q8blw4 mus musculu
39	270	12.9	1464	2	Q7TT32_MOUSE	Q7tt32 mus musculu
40	270	12.9	1467	2	Q5DTG2_MOUSE	Q5dtg2 mus musculu
41	268.5	12.8	742	2	Q9BYH7_HUMAN	Q9byh7 homo sapien
42	268.5	12.8	742	2	Q6P9F2_HUMAN	Q6p9f2 homo sapien
43	268.5	12.8	1258	2	Q8AW11_BRARE	Q8aw11 brachydanio
44	268.5	12.8	1835	2	Q9IAU4_CHICK	Q9iau4 gallus gall
45	267.5	12.8	316	2	Q19111_CAEEL	Q19111 caenorhabdi

ALIGNMENTS

RESULT 1
EDA_HUMAN
ID EDA_HUMAN STANDARD; PRT; 391 AA.
AC Q92838; O75910; Q5JUM7; Q9UP77; Q9Y6L0; Q9Y6L1; Q9Y6L2; Q9Y6L3;
AC Q9Y6L4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Ectodysplasin A (Ectodermal dysplasia protein) (EDA protein)
DE [Contains: Ectodysplasin A, membrane form; Ectodysplasin A, secreted
DE form].
GN Name=EDA; Synonyms=ED1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORM I), AND VARIANTS EDA
RP HIS-61 AND LEU-69.
RC TISSUE=Sweat gland;
RX MEDLINE=96331280; PubMed=8696334;
RA Kere J., Srivastava A.K., Montonen O., Zonana J., Thomas N.S.T.,
RA Ferguson B.M., Munoz F., Morgan D., Clarke A., Baybayan P., Chen E.Y.,
RA Ezer S., Saarialho-Kere U., la Chapelle A., Schlessinger D.;
RT "X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by
RT mutation in a novel transmembrane protein.";
RL Nat. Genet. 13:409-416(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM A1), AND VARIANTS EDA.
RC TISSUE=Liver;
RX MEDLINE=98349961; PubMed=9683615;
RA Monreal A.W., Zonana J., Ferguson B.M.;
RT "Identification of a new splice form of the EDA1 gene permits
RT detection of nearly all X-linked hypohidrotic ectodermal dysplasia
RT mutations.";
RL Am. J. Hum. Genet. 63:380-389(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE (ISOFORMS A1; A2; B; C; D; E AND F), AND VARIANTS
RP EDA.
RX MEDLINE=98409495; PubMed=9736768; DOI=10.1093/hmg/7.11.1661;
RA Bayes M., Hartung A.J., Ezer S., Pispa J., Thesleff I.,
RA Srivastava A.K., Kere J.;
RT "The anhidrotic ectodermal dysplasia gene (EDA) undergoes alternative
RT splicing and encodes ectodysplasin-A with deletion mutations in
RT collagenous repeats.";
RL Hum. Mol. Genet. 7:1661-1669(1998).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15772651; DOI=10.1038/nature03440;
RA Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K., Muzny D.,
RA Platzer M., Howell G.R., Burrows C., Bird C.P., Frankish A.,
RA Lovell F.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,
RA Jones M.C., Hurles M.E., Andrews T.D., Scott C.E., Searle S.,

RA Ramser J., Whittaker A., Deadman R., Carter N.P., Hunt S.E., Chen R.,
RA Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,
RA Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,
RA Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhya S.,
RA Ashwell R.I., Babbage A.K., Bagguley C.L., Ballabio A., Banerjee R.,
RA Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,
RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
RA Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W.,
RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
RA Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C.,
RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S.,
RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
RA Delgado O., Deshazo D., Dhami P., Ding Y., Dinh H., Dodsworth S.,
RA Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
RA Eades T., Ellwood M., Emery-Cohen A., Errington H., Evans K.L.,
RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P.,
RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
RA Griffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
RA Hawes A., Heath P.D., Heitmann K., Hennig S., Hernandez J.,
RA Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J.,
RA Hunt P.J., Hunt A.R., Isherwood J., Jacob L., Johnson D., Jones S.,
RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
RA Kioschis P., Klages S., Knights A.J., Kosiura A., Kovar-Smith C.,
RA Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W.,
RA Lloyd C., Lloyd D.M., Loulseged H., Loveland J.E., Lovell J.D.,
RA Lozado R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
RA McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T.,
RA Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Mueller I.,
RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,
RA Okwuonu G., Palmer S., Pandian R., Parker D., Parrish J.,
RA Pasternak S., Patel D., Pearce A.V., Pearson D.M., Pelan S.E.,
RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
RA Ridler K.A., Schlessinger D., Schueler M.G., Sehra H.K.,
RA Shaw-Smith C., Shen H., Sheridan E.M., Shownkeen R., Skuce C.D.,
RA Smith M.L., Sotheran E.C., Steingruber H.E., Steward C.A., Storey R.,
RA Swann R.M., Swarbreck D., Tabor P.E., Taudien S., Taylor T.,
RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
RA Tromans A.C., d'Urso M., Verduzco D., Villasana D., Waldron L.,
RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
RA Whitehead S.L., Whiteley M.N., Wilkinson J.E., Willey D.L.,
RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
RA Zorilla S., Buck D., Reinhardt R., Poustka A., Rosenthal A.,
RA Lehrach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
RA Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A.,
RA Nelson D.L., Weinstock G., Sulston J.E., Durbin R., Hubbard T.,
RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
RT "The DNA sequence of the human X chromosome.";
RL Nature 434:325-337(2005).
RN [5]
RP TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.
RA Kobielak K., Kobielak A., Trzciak W.H.;
RT "Expression of a novel transcript isoform of the EDA gene in human
RT umbilical cord.";
RL Eur. J. Hum. Genet. Suppl. 7:104-104(1999).
RN [6]
RP RECEPTOR INTERACTION (ISOFORMS A1 AND A2).
RX MEDLINE=20495245; PubMed=11039935; DOI=10.1126/science.290.5491.523;
RA Yan M., Wang L.-C., Hymowitz S.G., Schilbach S., Lee J., Goddard A.,
RA de Vos A.M., Gao W.-Q., Dixit V.M.;
RT "Two-amino acid molecular switch in an epithelial morphogen that
RT regulates binding to two distinct receptors.";
RL Science 290:523-527(2000).
RN [7]
RP PROCESSING, MUTAGENESIS OF ARG-153, AND CHARACTERIZATION OF VARIANT
RP HIS-156.
RX MEDLINE=21205766; PubMed=11309369; DOI=10.1093/hmg/10.9.953;
RA Elomaa O., Pulkkinen K., Hannelius U., Mikkola M., Saarialho-Kere U.,
RA Kere J.;
RT "Ectodysplasin is released by proteolytic shedding and binds to the
RT EDAR protein.";
RL Hum. Mol. Genet. 10:953-962(2001).
RN [8]
RP CHARACTERIZATION OF VARIANTS CYS-155; CYS-156 AND HIS-156, MUTAGENESIS
RP OF ARG-153; LYS-158 AND ARG-159, AND CLEAVAGE SITE.
RX MEDLINE=21309995; PubMed=11416205; DOI=10.1073/pnas.131076098;
RA Chen Y., Molloy S.S., Thomas L., Gambee J., Baechinger H.P.,
RA Ferguson B.M., Zonana J., Thomas G., Morris N.P.;
RT "Mutations within a furin consensus sequence block proteolytic release
RT of ectodysplasin-A and cause X-linked hypohidrotic ectodermal
RT dysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7218-7223(2001).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 230-391, AND SUBUNIT.
RX PubMed=14656435;
RA Hymowitz S.G., Compaan D.M., Yan M., Wallweber H.J., Dixit V.M.,
RA Starovasnik M.A., de Vos A.M.;
RT "The crystal structures of EDA-A1 and EDA-A2: splice variants with
RT distinct receptor specificity.";
RL Structure 11:1513-1520(2003).
RN [10]
RP VARIANT EDA TYR-54.
RX MEDLINE=98292028; PubMed=9630076;
RA Hertz J.M., Noergaard Hansen K., Juncker I., Kjeldsen M.,
RA Gregersen N.;
RT "A novel missense mutation (402C-->T) in exon 1 in the EDA gene in a
RT family with X-linked hypohidrotic ectodermal dysplasia.";
RL Clin. Genet. 53:205-209(1998).
RN [11]
RP VARIANT EDA LYS-63.
RX MEDLINE=98168231; PubMed=9507389;
RA Ferguson B.M., Thomas N.S.T., Munoz F., Morgan D., Clarke A.,
RA Zonana J.;
RT "Scarcity of mutations detected in families with X linked hypohidrotic
RT ectodermal dysplasia: diagnostic implications.";
RL J. Med. Genet. 35:112-115(1998).
RN [12]
RP VARIANT EDA ARG-55.
RX MEDLINE=99399307; PubMed=10469321;
RX DOI=10.1046/j.1523-1747.1999.00656.x;
RA Martinez F., Millan J.M., Orellana C., Prieto F.;
RT "X-linked anhidrotic (hypohidrotic) ectodermal dysplasia caused by a
RT novel mutation in EDA1 gene: 406T > G (Leu55Arg).";
RL J. Invest. Dermatol. 113:285-286(1999).
RN [13]
RP VARIANTS EDA ARG-60; TYR-252; VAL-269; SER-302 AND MET-378.
RX MEDLINE=21272350; PubMed=11378824; DOI=10.1038/sj/ejhg/5200635;
RA Vincent M.C., Biancalana V., Ginisty D., Mandel J.L., Calvas P.;
RT "Mutational spectrum of the ED1 gene in X-linked hypohidrotic
RT ectodermal dysplasia.";
RL Eur. J. Hum. Genet. 9:355-363(2001).
RN [14]
RP VARIANTS EDA CYS-156; HIS-156; CYS-255; ASP-255; GLY-274; TYR-332 AND
RP THR-349.
RX MEDLINE=21193173; PubMed=11295832; DOI=10.1002/humu.33;
RA Paeaekkoenen K., Cambiaghi S., Novelli G., Ouzts L.V., Penttinen M.,
RA Kere J., Srivastava A.K.;
RT "The mutation spectrum of the EDA gene in X-linked anhidrotic
RT ectodermal dysplasia.";
RL Hum. Mutat. 17:349-349(2001).
CC -!- FUNCTION: Seems to be involved in epithelial-mesenchymal signaling
CC during morphogenesis of ectodermal organs. Isoform A1 binds only
CC to the receptor EDAR, while isoform A2 binds exclusively to the
CC receptor XEDAR.
CC -!- SUBUNIT: Homotrimer. The homotrimers may then dimerize and form
CC higher order oligomers.
CC -!- INTERACTION:
CC Q9UNE0:EDAR; NbExp=1; IntAct=EBI-529425, EBI-529289;
CC -!- SUBCELLULAR LOCATION: Type II membrane protein and secreted.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=8;
CC Comment=Additional isoforms seem to exist;
CC Name=A1; Synonyms=II;
CC IsoId=Q92838-1; Sequence=Displayed;

```
CC       Name=I;
CC         IsoId=Q92838-2; Sequence=VSP_006454, VSP_006455;

  Query Match             100.0%;  Score 2098;  DB 1;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-115;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 2
EDA_BOVIN
ID   EDA_BOVIN      STANDARD;      PRT;   391 AA.
AC   Q9BEG5; Q9BEG6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Ectodysplasin A (Ectodysplasin 1) (Ectodermal dysplasia protein)
DE   [Contains: Ectodysplasin A, membrane form; Ectodysplasin A, secreted
DE   form].
GN   Name=EDA; Synonyms=ED1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC   Pecora; Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   NUCLEOTIDE SEQUENCE (ISOFORMS A1 AND A2).
RC   STRAIN=Holstein;
RX   MEDLINE=21070494; PubMed=11167539;
RA   Droegemueller C., Distl O., Leeb T.;
RT   "Identification of a highly polymorphic microsatellite within the
RT   bovine ectodysplasin A (ED1) gene on BTA Xq22-24.";
RL   Anim. Genet. 31:416-416(2000).
CC   -!- FUNCTION: Probably involved in epithelial-mesenchymal signaling.
CC       Isoform A1 binds only to the receptor EDAR, while isoform A2 binds
CC       exclusively to the receptor XEDAR (By similarity).
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein and secreted (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=A1;
CC         IsoId=Q9BEG5-1; Sequence=Displayed;
CC       Name=A2;
CC         IsoId=Q9BEG5-2; Sequence=VSP_006453;
CC   -!- PTM: N-glycosylated (By similarity).
CC   -!- PTM: Processing by furin produces a secreted form (By similarity).
CC   -!- DISEASE: Defects in EDA are the cause of anhidrotic ectodermal
CC       dysplasia. The disease is characterized by sparse hair (atrichosis
CC       or hypotrichosis), abnormal or missing teeth and the inability to
CC       sweat due to the absence of sweat glands.
CC   -!- SIMILARITY: Belongs to the tumor necrosis factor family.
CC   -!- SIMILARITY: Contains 1 collagen-like domain.
CC   ----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ----------------------------------------------------------------------
DR   EMBL; AJ300468; CAC29151.1; -; Genomic_DNA.
DR   EMBL; AJ300469; CAC29151.1; JOINED; Genomic_DNA.
DR   EMBL; AJ278907; CAC29151.1; JOINED; Genomic_DNA.
DR   EMBL; AJ300468; CAC29152.1; -; Genomic_DNA.
DR   EMBL; AJ300469; CAC29152.1; JOINED; Genomic_DNA.
DR   EMBL; AJ278907; CAC29152.1; JOINED; Genomic_DNA.
DR   SMR; Q9BEG5; 242-390.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR006052; TNF_family.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00229; TNF; 1.
DR   SMART; SM00207; TNF; 1.
DR   PROSITE; PS00251; TNF_1; FALSE_NEG.
DR   PROSITE; PS50049; TNF_2; 1.
KW   Alternative splicing; Collagen; Developmental protein;
KW   Differentiation; Glycoprotein; Signal-anchor; Transmembrane.
FT   CHAIN         1    391       Ectodysplasin A, membrane form.
FT   CHAIN       160    391       Ectodysplasin A, secreted form (By
FT                                similarity).
FT   TOPO_DOM      1     41       Cytoplasmic (Potential).
FT   TRANSMEM     42     62       Signal-anchor for type II membrane
FT                                protein (Potential).
FT   TOPO_DOM     63    391       Extracellular (Potential).
FT   DOMAIN      180    229       Collagen-like.
FT   SITE        159    160       Cleavage (by furin) (By similarity).
FT   CARBOHYD    313    313       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    372    372       N-linked (GlcNAc. . .) (Potential).
FT   VARSPLIC    307    308       Missing (in isoform A2).
FT                                /FTId=VSP_006453.
SQ   SEQUENCE   391 AA;  41567 MW;  1F87AD67A04EB7AA CRC64;

  Query Match             94.0%;  Score 1972;  DB 1;  Length 391;
  Best Local Similarity   94.4%;  Pred. No. 8.6e-108;
  Matches  369;  Conservative    4;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| |||||||||||||||| |||||||||||| ||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||||||| | |||||||||||| |||||: ||| | || ||:||||||||
Db         61 YLELRSELRRERGAESRFSGPGTPGTSGTLSSPGGLDPNGPITRHFGQRSPQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                | |||||||||:||| | || ||||||||||||||||:||||||||||||||||||
Db        121 TTLPPDSQDGHQMALVNFFIPKEKSYSEEESRRVRRNKRSKSSEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
```

```
Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 3
EDA_MOUSE
ID   EDA_MOUSE      STANDARD;      PRT;   391 AA.
AC   O54693; O35705; Q9QWJ8; Q9QZ01; Q9QZ02;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Ectodysplasin A (EDA protein homolog) (Tabby protein) [Contains:
DE   Ectodysplasin A, membrane form; Ectodysplasin A, secreted form].
GN   Name=Eda; Synonyms=Ed1, Ta;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE (ISOFORMS TAA; TAB AND TAC).
RC   STRAIN=129/Sv;
RX   MEDLINE=98058770; PubMed=9371801; DOI=10.1073/pnas.94.24.13069;
RA   Srivastava A.K., Pispa J., Hartung A.J., Du Y., Ezer S., Jenks T.,
RA   Shimada T., Pekkanen M., Mikkola M.L., Ko M.S.H., Thesleff I.,
RA   Kere J., Schlessinger D.;
RT   "The Tabby phenotype is caused by mutation in a mouse homologue of the
RT   EDA gene that reveals novel mouse and human exons and encodes a
RT   protein (ectodysplasin-A) with collagenous domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:13069-13074(1997).
RN   [2]
RP   NUCLEOTIDE SEQUENCE (ISOFORM TAD).
RX   MEDLINE=97449184; PubMed=9285798; DOI=10.1093/hmg/6.9.1589;
RA   Ferguson B.M., Brockdorff N., Formstone E., Ngyuen T.,
RA   Kronmiller J.E., Zonana J.;
RT   "Cloning of Tabby, the murine homolog of the human EDA gene: evidence
RT   for a membrane-associated protein with a short collagenous domain.";
RL   Hum. Mol. Genet. 6:1589-1594(1997).
RN   [3]
RP   NUCLEOTIDE SEQUENCE (ISOFORMS TA-A2 AND TA-A3).
RC   TISSUE=Embryo;
RX   MEDLINE=20005791; PubMed=10534613; DOI=10.1016/S0925-4773(99)00180-X;
RA   Mikkola M.L., Pispa J., Pekkanen M., Paulin L., Nieminen P., Kere J.,
RA   Thesleff I.;
RT   "Ectodysplasin, a protein required for epithelial morphogenesis, is a
RT   novel TNF homologue and promotes cell-matrix adhesion.";
RL   Mech. Dev. 88:133-146(1999).
CC   -!- FUNCTION: Involved in epithelial-mesenchymal signaling during
CC       morphogenesis of ectodermal organs. Isoform TAA binds only to the
CC       receptor EDAR, while isoform TA-A2 binds exclusively to the
CC       receptor XEDAR.
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein and secreted (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC         Comment=Additional isoforms seem to exist;
CC       Name=TAA; Synonyms=A1;
CC         IsoId=O54693-1; Sequence=Displayed;
CC       Name=TA-A2;
CC         IsoId=O54693-2; Sequence=VSP_006471;
CC       Name=TA-A3;
CC         IsoId=O54693-3; Sequence=VSP_006469, VSP_006471;
CC       Name=TAB;
CC         IsoId=O54693-4; Sequence=VSP_006466, VSP_006467;
CC       Name=TAC;
CC         IsoId=O54693-5; Sequence=VSP_006465, VSP_006468;
CC       Name=TAD;
CC         IsoId=O54693-6; Sequence=VSP_006470;
CC   -!- PTM: N-glycosylated (By similarity).
CC   -!- PTM: Processing by furin produces a secreted form (By similarity).
CC   -!- DISEASE: Defects in Eda are the cause of the tabby phenotype in
CC       mice (the equivalent of anhidrotic ectodermal dysplasia in
CC       humans). The disease is characterized by sparse hair (atrichosis
CC       or hypotrichosis), abnormal or missing teeth and the inability to
CC       sweat due to the absence of sweat glands.
CC   -!- SIMILARITY: Belongs to the tumor necrosis factor family.
CC   -!- SIMILARITY: Contains 1 collagen-like domain.
CC   ----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ----------------------------------------------------------------------
DR   EMBL; AF016627; AAB95202.1; -; mRNA.
DR   EMBL; AF016628; AAB95203.1; -; mRNA.
DR   EMBL; AF016629; AAB95204.1; -; mRNA.
DR   EMBL; AF016630; AAB95205.1; -; mRNA.
DR   EMBL; AF016631; AAB95206.1; -; mRNA.
DR   EMBL; AF004434; AAB88121.1; -; Genomic_DNA.
DR   EMBL; AF004435; AAB88122.1; -; mRNA.
DR   EMBL; Y13438; CAA73849.1; -; Genomic_DNA.
DR   EMBL; AJ243657; CAB52696.1; -; mRNA.
DR   EMBL; AJ243658; CAB52697.1; -; mRNA.
DR   SMR; O54693; 242-390.
DR   Ensembl; ENSMUSG00000059327; Mus musculus.
DR   MGI; MGI:1195272; Eda.
DR   GO; GO:0045177; C:apical part of cell; IDA.
DR   GO; GO:0005789; C:endoplasmic reticulum membrane; IDA.
DR   GO; GO:0005887; C:integral to plasma membrane; IDA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IDA.
DR   GO; GO:0042346; P:positive regulation of NF-kappaB-nucleus im. . .; IDA.
DR   GO; GO:0007431; P:salivary gland development; IDA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR006052; TNF_family.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00229; TNF; 1.
DR   SMART; SM00207; TNF; 1.
DR   PROSITE; PS00251; TNF_1; FALSE_NEG.
DR   PROSITE; PS50049; TNF_2; 1.
KW   Alternative splicing; Collagen; Developmental protein;
KW   Differentiation; Glycoprotein; Signal-anchor; Transmembrane.
FT   CHAIN         1    391       Ectodysplasin A, membrane form.
FT   CHAIN       160    391       Ectodysplasin A, secreted form (By
FT                                similarity).
FT   TOPO_DOM      1     41       Cytoplasmic (Potential).
FT   TRANSMEM     42     62       Signal-anchor for type II membrane
FT                                protein (Potential).
FT   TOPO_DOM     63    391       Extracellular (Potential).
FT   DOMAIN      180    229       Collagen-like.
FT   SITE        159    160       Cleavage (by furin) (By similarity).
FT   CARBOHYD    313    313       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    372    372       N-linked (GlcNAc. . .) (Potential).
FT   VARSPLIC    133    238       MALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNK
FT                                KKGKKAGPPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVM
FT                                GPPGPPGPPGPQGPPGLQGPSGAA -> VSHLGGAAALEAP
FT                                SPARLGGGLGLRAQGTLPLRAKFQGRSWEWAGVLGRGCPGQ
FT                                VVLGSCLGSSRPSPVPWSWKAQPARAAPGEVWAA (in
FT                                isoform TAC).
FT                                /FTId=VSP_006465.
FT   VARSPLIC    169    177       PVKNKKKGK -> KSTQVIFFP (in isoform TAB).
FT                                /FTId=VSP_006466.
FT   VARSPLIC    178    391       Missing (In isoform TAB).
FT                                /FTId=VSP_006467.
FT   VARSPLIC    239    391       Missing (In isoform TAC).
FT                                /FTId=VSP_006468.
FT   VARSPLIC    265    267       Missing (In isoform TA-A3).
FT                                /FTId=VSP_006469.
FT   VARSPLIC    295    308       Missing (In isoform TAD).
FT                                /FTId=VSP_006470.
FT   VARSPLIC    307    308       Missing (In isoform TA-A2 and isoform TA-
FT                                A3).
FT                                /FTId=VSP_006471.
```

```
FT   CONFLICT    126    126       D -> E (in Ref. 2).
SQ   SEQUENCE   391 AA;  41603 MW;  E5ECEDA5BD60DBFF CRC64;

  Query Match             93.7%;  Score 1965;  DB 1;  Length 391;
  Best Local Similarity   94.6%;  Pred. No. 2.2e-107;
  Matches  370;  Conservative    1;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| |||||||||||||||| |||||||||| ||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||| |||||| | |||||||| | |||  ||| |||||| :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              |  |||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db        121 DPLPPDSQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 4
Q58P77_CANFA
ID   Q58P77_CANFA PRELIMINARY;      PRT;   386 AA.
AC   Q58P77;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Ectodysplasin A isoform EDA-A2.
GN   Name=EDA;
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC   Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Casal M.L., Scheidt J.L., Rhodes J.L., Henthorn P.S., Werner P.;
RT   "Mutation identification in a canine model of X-linked ectodermal
RT   dysplasia.";
RL   Submitted (FEB-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY924414; AAX38257.1; -; Genomic_DNA.
DR   EMBL; AY924407; AAX38257.1; JOINED; Genomic_DNA.
DR   EMBL; AY924408; AAX38257.1; JOINED; Genomic_DNA.
DR   EMBL; AY924410; AAX38257.1; JOINED; Genomic_DNA.
DR   EMBL; AY924409; AAX38257.1; JOINED; Genomic_DNA.
DR   EMBL; AY924413; AAX38257.1; JOINED; Genomic_DNA.
DR   EMBL; AY924412; AAX38257.1; JOINED; Genomic_DNA.
DR   EMBL; AY924411; AAX38257.1; JOINED; Genomic_DNA.
DR   SMR; Q58P77; 237-385.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005164; F:tumor necrosis factor receptor binding; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR006052; TNF_family.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00229; TNF; 1.
DR   PROSITE; PS50049; TNF_2; 1.
KW   Collagen.
SQ   SEQUENCE   386 AA;  41235 MW;  6095BDBC3EACF68B CRC64;

  Query Match             90.1%;  Score 1890.5;  DB 2;  Length 386;
  Best Local Similarity   91.3%;  Pred. No. 5e-103;
  Matches  357;  Conservative    8;  Mismatches   21;  Indels    5;  Gaps    2;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| || |||||||||||||| ||||:||||||| ||||||||||||||||||
Db          1 MGYPEVERREPLPTAAPRERGSQGCGCRGAPAQAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||    ||| |||:| |||||| |||   |||||:|||||  |
Db         61 YLELRSELRRERGAESRLG----PGTPGTLNSPGGLDPDGPITRDSGQPSPQQQPLEAEE 116

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||   |:||||||||||||:|| |||:| || |||||||::|| ||||||||||||||||
Db        117 TALPPHSRDGHQMALLNFFFPEEKSYSEDE-RRFRRNKRSKNSEGTDGPVKNKKKGKKAG 175

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        176 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGTADK 235

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 AGPRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 295

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 355

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        356 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 386


RESULT 5
Q5EFZ5_CHICK
ID   Q5EFZ5_CHICK PRELIMINARY;      PRT;   356 AA.
AC   Q5EFZ5;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Ectodysplasin A1 (Fragment).
GN   Name=Eda;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Houghton L., Morgan B.A.;
RL   Submitted (JAN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY885699; AAW81692.1; -; mRNA.
DR   SMR; Q5EFZ5; 207-355.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005164; F:tumor necrosis factor receptor binding; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR006052; TNF_family.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00229; TNF; 1.
DR   PROSITE; PS50049; TNF_2; 1.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   356 AA;  37526 MW;  8D53FD97F71FDDFB CRC64;

  Query Match             68.7%;  Score 1441.5;  DB 2;  Length 356;
```

```
  Best Local Similarity   74.9%;  Pred. No. 9.2e-77;
  Matches  289;  Conservative   19;  Mismatches   41;  Indels   37;  Gaps    9;

Qy         13 PAAAPRERGSQ----GCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |||   ||  :    || | ||        |  |||||| ||||||:||| |||||||||
Db          1 PAAMGAERRGEAARGGCSCAGA-------GSWRLFLGFFALSLALHVLTLGCYLELRSEL 53

Qy         69 RRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPK-QQPLEPGEAALHSDS 127
              ||:|| :     :  |   || ::     | :|       |: : |:| | ||
Db         54 RRDRGPQP----AAPPRRDGTAAAAA---PGAP-------PAVRPQRPAESGERR----- 94

Qy        128 QDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGP--VKNKKKGKKAGPPGPN 185
                 |:|||||| |:|| :   | |||||||||| :|| |||  ||||||||||||||||
Db         95 ---QQLALLNFFHPEEKLH-VGEGRRVRRNKRSKGSEGPDGPSSVKNKKKGKKAGPPGPN 150

Qy        186 GPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRE 245
              || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||:|:
Db        151 GPQGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGATDKAGSRD 210

Qy        246 NQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYS 305
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        211 TQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHARSGELEVLVDGTYFIYS 270

Qy        306 QVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKIAVKMV 365
              |||||||||||||||||||||||||||||| |||:||||||||||||||||||||||||
Db        271 QVEVYYINFTDFASYEVVVDEKPFLQCTRSIETSKTNFNTCYTAGVCLLKARQKIAVKMV 330

Qy        366 HADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||:||||
Db        331 HADISINMSKHTTFFGAIRLGDAPAS 356
```

RESULT 6
Q50D54_GASAC
ID   Q50D54_GASAC PRELIMINARY;   PRT;   367 AA.
AC   Q50D54;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Ectodysplasin.
GN   Name=Eda;
OS   Gasterosteus aculeatus (Three-spined stickleback).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
OC   Gasterosteidae; Gasterosteus.
OX   NCBI_TaxID=69293;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15790847; DOI=10.1126/science.1107239;
RA   Colosimo P.F., Hosemann K.E., Balabhadra S., Villarreal G. Jr,
RA   Dickson M., Grimwood J., Schmutz J., Myers R.M., Schluter D.,
RA   Kingsley D.M.;
RT   "Widespread parallel evolution in sticklebacks by repeated fixation of
RT   Ectodysplasin alleles.";
RL   Science 307:1928-1933(2005).
DR   EMBL; AY897589; AAY27076.1; -; Genomic_DNA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR006052; TNF_family.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00229; TNF; 1.
DR   PROSITE; PS00251; TNF_1; UNKNOWN_1.
DR   PROSITE; PS50049; TNF_2; 1.
KW   Collagen.
SQ   SEQUENCE   367 AA; 39746 MW; C0BFEC40D53A9EF6 CRC64;

```
  Query Match             45.5%;  Score 955.5;  DB 2;  Length 367;
  Best Local Similarity   51.1%;  Pred. No. 2.8e-48;
  Matches  192;  Conservative   59;  Mismatches   86;  Indels   39;  Gaps    9;

Qy         25 CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRERGAESRLGGSGTP 84
              | |      |:  |:  ::||| | |||:|| :|| |||:||||::||   : |
Db         19 CTC-NKKCRSRSGS--VVFLGLFLLSLSLHAVTLVCYLDLRSEVKREIIHQKR------- 68

Qy         85 GTSGTLSSLGGLDPDSPITSHL-GQPSPKQQPLEPGEAALHSDSQDGHQMALLNFFFPDE 143
                  |: :  |:||  |      ||| |       ||         ||:  ||:    :
Db         69 ---DTMLTFAGIDPADPAAVLAPGQPRPDSGVGRGGE---------GHEEKLLH---RNG 113

Qy        144 KPYSEEESRRVRRNKRSKSNEGADGPV--------KNKKKGKKAGPPGPNGPPGPPGPPG 195
              : :: :::| :   :|:| : |               | ||||||   |   ||||||||||
Db        114 ELHATQDNRGI--TQRAKRSPGKQPETELTGRERRKEKKKGKKRPVP---GPPGPPGPPG 168

Qy        196 PQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVVHLQG 255
              |||||||||||||||:  :|| ||||||||||||  ||:|  ||  |:| |||||||||
Db        169 PQGPPGIPGIPGIPGSNAVGPAGPPGPPGPQGPPGTPGPAGVPDKTKTKEFQPAVVHLQG 228

Qy        256 QGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFT 315
              | : |||: ||| |:| :|  :::: :|||:| |||||||:|| ||||||||||:|||
Db        229 QETTIQVREDLSEGILRNWKMVSIHHRVFKMHSRSGELEVLLDGVYFIYSQVEVYYLNFT 288

Qy        316 DFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSK 375
              | |||||:||  |||:|| |||||:  :|||||||| ||:| |:|::::|: |  |:|:
Db        289 DIASYEVMVDSNPFLRCTCSIETGQRKFNTCYTAGVSLLRAGQRISIRIVYEDTLISMTN 348

Qy        376 HTTFFGAIRLGEAPAS 391
              |||| |::||||||::
Db        349 HTTFLGSVRLGEAPSA 364
```

RESULT 7
Q5JS00_HUMAN
ID   Q5JS00_HUMAN PRELIMINARY;   PRT;   178 AA.
AC   Q5JS00;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Ectodysplasin A.
GN   Name=EDA; ORFNames=RP11-351K23.1-003;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Whitehead S.;
RL   Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   Clark S.;
RL   Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL450449; CAI39805.1; -; Genomic_DNA.
DR   EMBL; AL158069; CAI41611.1; -; Genomic_DNA.
DR   EMBL; AL158069; CAI39805.1; JOINED; Genomic_DNA.
DR   EMBL; AL450449; CAI41611.1; JOINED; Genomic_DNA.
SQ   SEQUENCE   178 AA; 19200 MW; A390EF5B3F6B865A CRC64;

```
  Query Match             42.3%;  Score 887;  DB 2;  Length 178;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-44;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADG 168
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADG 168
```

RESULT 8
Q4SJR1_TETNG

```
ID   Q4SJR1_TETNG PRELIMINARY;      PRT;   207 AA.
AC   Q4SJR1;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Chromosome 1 SCAF14573, whole genome shotgun sequence.
DE   (Fragment).
GN   ORFNames=GSTENG00017090001;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA   Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA   Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA   Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA   Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA   Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA   Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA   Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA   Kellis M., Volff JN., Guigo R., Zody M.C., Mesirov J.,
RA   Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA   Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA   Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT   "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT   the early vertebrate proto-karyotype.";
RL   Nature 431:946-957(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RG   Genoscope; Whitehead Institute Centre for Genome Research;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAE01014573; CAF99121.1; -; Genomic_DNA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR006052; TNF_family.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00229; TNF; 1.
DR   PROSITE; PS00251; TNF_1; UNKNOWN_1.
DR   PROSITE; PS50049; TNF_2; 1.
KW   Collagen.
FT   NON_TER       1      1
FT   NON_TER     207    207
SQ   SEQUENCE   207 AA;  22149 MW;  6CB874AF177DA5CA CRC64;

  Query Match           38.6%;  Score 809;  DB 2;  Length 207;
  Best Local Similarity   68.9%;  Pred. No. 6e-40;
  Matches  142;  Conservative   32;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        186 GPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRE 245
              ||||||||||||||||||||||||||: |:|| ||||||||||||| |||:|  ||  |||
Db          1 GPPGPPGPPGPQGPPGIPGIPGIPGSNVVGPVGPPGPPGPQGPPGTQGPAGDPDKTKTRE 60

Qy        246 NQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYS 305
              | |||||||| : |||: ||| |:| :|  :::: :|||:| |||||||:|| |||||
Db         61 FQAAVVHLQGQETTIQVREDLSEGILRNWKMVSIHHRVFKMHSRSGELEVLLDGVYFIYS 120

Qy        306 QVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKIAVKMV 365
              |||||:|||| |||||:||  |||:|| |||||:  :|||||||| ||:| |:|::::|
Db        121 QVEVYYLNFTDIASYEVMVDSNPFLRCTCSIETGQRKFNTCYTAGVSLLRAGQRISIRIV 180

Qy        366 HADISINMSKHTTFFGAIRLGEAPAS 391
              : |  |:|: |||| |::||||||::
Db        181 YEDTLISMTNHTTFLGSVRLGEAPSA 206


RESULT 9
Q8VIF6_MOUSE
ID   Q8VIF6_MOUSE PRELIMINARY;      PRT;   742 AA.
AC   Q8VIF6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Scavenger receptor with C-type lectin.
GN   Name=Colec12; Synonyms=srcl;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=21575692; PubMed=11718900; DOI=10.1016/S0167-4781(01)00284-6;
RA   Nakamura K., Funakoshi H., Tokunaga F., Nakamura T.;
RT   "Molecular cloning of a mouse scavenger receptor with C-type lectin
RT   (SRCL)(1), a novel member of the scavenger receptor family.";
RL   Biochim. Biophys. Acta 1522:53-58(2001).
DR   EMBL; AB038519; BAB82497.1; -; mRNA.
DR   HSSP; P22897; 1EGI.
DR   MGI; MGI:2152907; Colec12.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0016020; C:membrane; ISS.
DR   GO; GO:0005534; F:galactose binding; ISS.
DR   GO; GO:0030169; F:low-density lipoprotein binding; ISS.
DR   GO; GO:0008329; F:pattern recognition receptor activity; ISS.
DR   GO; GO:0005044; F:scavenger receptor activity; ISS.
DR   GO; GO:0019733; P:antibacterial humoral response (sensu Verte. . .; ISS.
DR   GO; GO:0019734; P:antifungal humoral response (sensu Vertebrata); ISS.
DR   GO; GO:0009756; P:carbohydrate mediated signaling; ISS.
DR   GO; GO:0045087; P:innate immune response; ISS.
DR   GO; GO:0045194; P:oxidized low-density lipoprotein catabolism; ISS.
DR   GO; GO:0006910; P:phagocytosis, recognition; ISS.
DR   GO; GO:0051260; P:protein homooligomerization; ISS.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; Lectin_C; 1.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Lectin; Receptor.
SQ   SEQUENCE   742 AA;  81306 MW;  85A90D3AE881DB6B CRC64;

  Query Match           13.6%;  Score 286;  DB 2;  Length 742;
  Best Local Similarity   30.5%;  Pred. No. 9.2e-09;
  Matches   84;  Conservative   29;  Mismatches   96;  Indels   66;  Gaps   11;

Qy         82 GTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNFFFP 141
              | ||  |      |   |  | | : ||   | :|  || |         | :  :     |
Db        443 GPPGPRGPKGDRGSQGPPGP-TGNKGQKGEKGEPGPPGPA--------GERGTIGPVGPP 493

Qy        142 DEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGP------PG 195
              |          : ||.::|: ||  ::    | || ||:| ||||||       ||
Db        494 GE--------------RGSKGSKGSQGPKGSRGSPGKPGPQGPSGDPGPPGPPGKDGLPG 539

Qy        196 PQGPPGIPGI------PGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQP 248
              ||||||  |:      ||:||   : | || || |||:|||| ||||| :     :|:|
Db        540 PQGPPGFQGLQGTVGEPGVPGPRGLPGLPGVPGMPGPKGPPGPPGPSGAMEPLAL-QNEP 598

Qy        249 AVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVE 308
                    | :|      |    |    |      | |      | |:| |   |
Db        599 T--------PASEV-----NGCPPHWKNFTDKCYYFSL-----EKEILEDAKLFC----- 635

Qy        309 VYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNY 343
                    | :|: | :: :   |  :    |: ::
Db        636 ------EDKSSHLVFINSREEQQWIKKHTVGRESH 664


RESULT 10
```

```
Q8CFM4_MOUSE
ID   Q8CFM4_MOUSE PRELIMINARY;      PRT;   998 AA.
AC   Q8CFM4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Col3a1 protein (Fragment).
GN   Name=Col3a1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Mix FVB/N;
RC   TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Mix FVB/N;
RC   TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC013626; AAH13626.1; -; mRNA.
DR   Ensembl; ENSMUSG00000026043; Mus musculus.
DR   MGI; MGI:88453; Col3a1.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 12.
DR   ProDom; PD000007; Clg_helix; 2.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   998 AA;  95451 MW;  2AB4BA953B7084A6 CRC64;

  Query Match             13.5%;  Score 284;  DB 2;  Length 998;
  Best Local Similarity   30.3%;  Pred. No. 1.6e-08;
  Matches   87;  Conservative   15;  Mismatches   91;  Indels   94;  Gaps   11;

Qy         13 PAAAPRERGSQG----CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |  | |||  |     |  ||| : ||  :
Db        548 PRGGPGERGEHGPPGPAGFPGAPGQNGEPGA---------------------------- 578

Qy         69 RRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              : ||||    |  | || :|   | |   |  |     |    :  |  || |
Db        579 KGERGAPGEKGEGGPPGPAGPTGSSGPAGPPGP----QGVKGERGSPGGPGTAG------ 628

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
                          ||                    |  ||  |       |||||:| |
Db        629 -----------FP--------------------GGRGLPGPPGN---NGNPGPPGPSGAP 654

Qy        189 GPPGPPGPQGPPGIPGIPGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR--E 245
              |  ||||| |  | || ||| |      | ||  |||| |||||  || | |   | |
Db        655 GKDGPPGPAGNSGSPGNPGIAGPKGDAGQPGEKGPPGAQGPPGSPGPLGIAGLTGARGLA 714

Qy        246 NQPAV----VHLQGQGSAIQVKND----LSGGVLNDWSRITMNPKVF 284
                 | :    | |  : | ::|:|     : | |       : ::||||
Db        715 GPPGMPGPRVALDLRVSRVKVENQEPVAIMGNV------VLLDPKVF 755


RESULT 11
Q8K4Q8_MOUSE
ID   Q8K4Q8_MOUSE PRELIMINARY;      PRT;   742 AA.
AC   Q8K4Q8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Collectin placenta 1 (Collectin sub-family member 12).
GN   Name=Colec12; Synonyms=CL-P1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Keshi H., Sakai Y.,
RA   Fukuoh A., Sakamoto T., Itabe H., Suzutani T., Ogasawara M.,
RA   Yoshida I., Wakamiya N.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;
RA   Director MGC Project;
RL   Submitted (SEP-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB078434; BAC05523.1; -; mRNA.
DR   EMBL; BC057936; AAH57936.1; -; mRNA.
DR   HSSP; P22897; 1EGI.
DR   Ensembl; ENSMUSG00000036103; Mus musculus.
DR   MGI; MGI:2152907; Colec12.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0016020; C:membrane; ISS.
DR   GO; GO:0005534; F:galactose binding; ISS.
DR   GO; GO:0030169; F:low-density lipoprotein binding; ISS.
DR   GO; GO:0008329; F:pattern recognition receptor activity; ISS.
DR   GO; GO:0005044; F:scavenger receptor activity; ISS.
DR   GO; GO:0019733; P:antibacterial humoral response (sensu Verte. . .; ISS.
DR   GO; GO:0019734; P:antifungal humoral response (sensu Vertebrata); ISS.
DR   GO; GO:0009756; P:carbohydrate mediated signaling; ISS.
DR   GO; GO:0045087; P:innate immune response; ISS.
```

```
DR   GO; GO:0045194; P:oxidized low-density lipoprotein catabolism; ISS.
DR   GO; GO:0006910; P:phagocytosis, recognition; ISS.
DR   GO; GO:0051260; P:protein homooligomerization; ISS.
DR   GO; GO:0044404; P:symbiotic interaction between host and othe. . .; ISS.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; Lectin_C; 1.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   742 AA;  81303 MW;  1537C490E5911C45 CRC64;

  Query Match             13.4%;  Score 282;  DB 2;  Length 742;
  Best Local Similarity   30.2%;  Pred. No. 1.6e-08;
  Matches   83;  Conservative   29;  Mismatches   97;  Indels   66;  Gaps   11;

Qy         82 GTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNFFFP 141
              | || |      |   | | | : ||   | :|  || |         | :  :     |
Db        443 GPPGPRGPKGDRGSQGPPGP-TGNKGQKGEKGEPGPPGPA--------GERGTIGPVGPP 493

Qy        142 DEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGP------PG 195
               |              : || ::|: ||  ::    | || ||:| ||||||      ||
Db        494 GE-------------RGSKGSKGSQGPKGSRGSPGKPGPQGPSGDPGPPGPPGKDGLPG 539

Qy        196 PQGPPGIPGI------PGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQP 248
              ||||||  |:      ||:||   : | || || |||:|||| |||||  :     :|:|
Db        540 PQGPPGFQGLQGTVGEPGVPGPRGLPGLPGVPGMPGPKGPPGPPGPSGAMEPLAL-QNEP 598

Qy        249 AVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVE 308
                  | :|      |    |   |     | |      | |:  |   |
Db        599 T--------PASEV-----NGCPPHWKNFTDKCYYPSL-----EKEIFEDAKLFC----- 635

Qy        309 VYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNY 343
                  | :|: | :: :   |  :    |: ::
Db        636 ------EDKSSHLVFINSREEQQWIKKHTVGRESH 664


RESULT 12
Q8C979_MOUSE
ID   Q8C979_MOUSE PRELIMINARY;      PRT;   742 AA.
AC   Q8C979;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
DE   enriched library, clone:A730023E20 product:collectin sub-family member
DE   12, full insert sequence.
GN   Name=Colec12;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK042772; BAC31361.1; -; mRNA.
DR   HSSP; P22897; 1EGI.
DR   MGI; MGI:2152907; Colec12.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; Lectin_C; 1.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
```

```
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   742 AA;  81289 MW;  1350DCCB34522C45 CRC64;

  Query Match           13.4%;  Score 282;  DB 2;  Length 742;
  Best Local Similarity   30.2%;  Pred. No. 1.6e-08;
  Matches   83;  Conservative   29;  Mismatches   97;  Indels   66;  Gaps   11;

Qy         82 GTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNFFFP 141
              | ||  |      |   | | | : ||   | :|  || |         | :  :     |
Db        443 GPPGPRGPKGDRGSQGPPGP-TGNKGQKGEKGEPGPPGPA--------GERGTIGPVGPP 493

Qy        142 DEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGP------PG 195
                |           : || ::|: ||  ::    | || ||:| ||||||      ||
Db        494 GE-------------RGSKGSKGSQGPKGSRGSPGKPGPQGPSGDPGPPGPPGKDGLPG 539

Qy        196 PQGPPGIPGI------PGIPGTT-VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQP 248
              ||||||  |:      ||:||   : | || || |||:|||| |||||  :     :|:|
Db        540 PQGPPGFQGLQGTVGEPGVPGPRGLPGLPGVPGMPGPKGPPGPPGPSGAMEPLAL-QNEP 598

Qy        249 AVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVE 308
                  | :|      |   |   |     | |   | |:  |   |
Db        599 T--------PASEV-----NGCPPHWKNFTDKCYYFSL-----EKEIFEDAKLFC----- 635

Qy        309 VYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNY 343
                  | :|: | :: :   |  :    |: ::
Db        636 ------EDKSSHLVFINSREEQQWIKKHTVGRESH 664


RESULT 13
Q59EE7_HUMAN
ID   Q59EE7_HUMAN PRELIMINARY;      PRT;  1792 AA.
AC   Q59EE7;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Pro-alpha-1 type V collagen variant (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Aorta endothelial cell;
RA   Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,
RA   Ohara O., Nagase T., Kikuno F.R.;
RT   "None Title.";
RL   Submitted (MAR-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB209864; BAD93101.1; -; mRNA.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; Laminin_G_TSP_N.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen; Extracellular matrix.
FT   NON_TER       1       1
SQ   SEQUENCE   1792 AA;  178517 MW;  E47B3093C34C5A9D CRC64;

  Query Match           13.4%;  Score 280.5;  DB 2;  Length 1792;
  Best Local Similarity   30.1%;  Pred. No. 4.7e-08;
  Matches   83;  Conservative    9;  Mismatches   69;  Indels  115;  Gaps   10;

Qy         13 PAAAPRERGSQGC-GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRE 71
                 |   | |:|  |  |  | | | |   |    :|| |

Db        811 PLGPPGEKGKLGVPGLPGYPGRQGPKGS----IGFPGFP---------------------- 845

Qy         72 RGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGH 131
                ||    || |||                          |:| |: |    |
Db        846 -GANGEKGGRGTP---------------------GKPGPRGQRGPTG------------ 870

Qy        132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPP---GPNGPP 188
                          |  |   | :     | | | |||          ||||   |||||
Db        871 -------------PRGERGPRGITGKPGPKGNSGGDGP---------AGPPGERGPNGPQ 908

Qy        189 GPPGPPGPQGPPGIPGIPGIPG-------TTVMGPPGPPGP------------------- 222
                || | |||:|||| ||  |:||       |   |  |||||
Db        909 GPTGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGER 968

Qy        223 --PGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQ 256
                ||| |||| ||  | | | ||:   |    | |:
Db        969 GHPGPPGPPGEQGLPGLAGKEGTK-GDPGPAGLPGK 1003


RESULT 14
CO5A1_HUMAN
ID   CO5A1_HUMAN    STANDARD;      PRT;  1838 AA.
AC   P20908;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Collagen alpha 1(V) chain precursor.
GN   Name=COL5A1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE, AND PROTEIN SEQUENCE OF 556-565.
RX   MEDLINE=91302336; PubMed=2071595;
RA   Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,
RA   Kato I.;
RT   "Complete primary structure of human collagen alpha 1 (V) chain.";
RL   J. Biol. Chem. 266:13124-13129(1991).
RN   [2]
RP   PROTEIN SEQUENCE OF 621-822.
RC   TISSUE=Chorioamniotic membrane;
RX   MEDLINE=89227189; PubMed=2496661;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of three cyanogen
RT   bromide-derived peptides from human alpha 1(V) collagen chain.";
RL   Arch. Biochem. Biophys. 271:120-129(1989).
RN   [3]
RP   PROTEIN SEQUENCE OF 823-950, AND HEPARIN-BINDING.
RX   MEDLINE=90366601; PubMed=2203476; DOI=10.1016/0304-4165(90)90108-9;
RA   Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;
RT   "Primary structure of the heparin-binding site of type V collagen.";
RL   Biochim. Biophys. Acta 1035:139-145(1990).
RN   [4]
RP   PROTEIN SEQUENCE OF 556-571.
RC   TISSUE=Placenta;
RX   MEDLINE=92239022; PubMed=1571108;
RA   Mann K.;
RT   "Isolation of the alpha 3-chain of human type V collagen and
RT   characterization by partial sequencing.";
RL   Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN   [5]
RP   PROTEIN SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND
RP   1465-1477.
RC   TISSUE=Chorioamniotic membrane;
RX   MEDLINE=94237164; PubMed=8181482;
RA   Moradi-Ameli M., Rousseau J.C., Kleman J.P., Champliaud M.F.,
RA   Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT   "Diversity in the processing events at the N-terminus of type-V
RT   collagen.";
RL   Eur. J. Biochem. 221:987-995(1994).
RN   [6]
```

```
RP   DISEASE, AND VARIANT EDS-I SER-1639.
RX   MEDLINE=97195540; PubMed=9042913;
RA   de Paepe A., Nuytinck L., Hausser I., Anton-Lamprecht I.,
RA   Naeyaert J.-M.;
RT   "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos
RT   syndromes I and II.";
RL   Am. J. Hum. Genet. 60:547-554(1997).
RN   [7]
RP   VARIANTS EDS-1 SER-530 AND ASP-1489.
RX   MEDLINE=20068401; PubMed=10602121;
RX   DOI=10.1002/(SICI)1096-8628(20000103)90:1<72::AID-AJMG13>3.0.CO;2-C;
RA   Giunta C., Steinmann B.;
RT   "Compound heterozygosity for a disease-causing G1489D and disease-
RT   modifying G530S substitution in COL5A1 of a patient with the classical
RT   type of Ehlers-Danlos syndrome: an explanation of intrafamilial
RT   variability?";
RL   Am. J. Med. Genet. 90:72-79(2000).
CC   -!- FUNCTION: Type V collagen is a member of group I collagen
CC       (fibrillar forming collagen). It is a minor connective tissue
CC       component of nearly ubiquitous distribution. Type V collagen binds
CC       to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC   -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
CC       most tissues and trimers of one alpha 1(V), one alpha 2(V), and
CC       one alpha 3(V) chains in placenta.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- PTM: Sulfated on 40% of tyrosines.
CC   -!- DISEASE: Defects in COL5A1 are a cause of Ehlers-Danlos syndrome
CC       type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
CC       gravis. EDS-I is a connective-tissue disorder characterized by
CC       loose-jointedness and fragile, velvety, stretchable, bruisable
CC       skin that heals with peculiar 'cigarette-paper' scars. Inheritance
CC       is autosomal dominant.
CC   -!- DISEASE: Defects in COL5A1 are a cause of Ehlers-Danlos syndrome
CC       type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
CC       syndrome mitis. Inheritance is autosomal dominant.
CC   -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; D90279; BAA14323.1; -; mRNA.
DR   PDB; 1A89; Model; A/B/C=904-924.
DR   PDB; 1A9A; Model; A/C=904-924.
DR   Ensembl; ENSG00000130635; Homo sapiens.
DR   HGNC; HGNC:2209; COL5A1.
DR   MIM; 120215; -.
DR   MIM; 130000; -.
DR   MIM; 130010; -.
DR   GO; GO:0005588; C:collagen type V; TAS.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; Laminin_G_TSP_N.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   3D-structure; Collagen; Direct protein sequencing; Disease mutation;
KW   Ehlers-Danlos syndrome; Extracellular matrix; Heparin-binding;
KW   Hydroxylation; Repeat; Signal; Structural protein; Sulfation.
FT   SIGNAL       1     37       Potential.
FT   CHAIN       38   1605       Collagen alpha 1(V) chain.
FT   PROPEP    1606   1838       C-terminal propeptide.
FT   DOMAIN      39    230       TSP N-terminal.
FT   REGION     231    443       Nonhelical region.
FT   REGION     444    558       Interrupted collagenous region.
FT   REGION     559   1570       Triple-helical region.
FT   REGION    1571   1605       Nonhelical region.
FT   MOD_RES    234    234       Sulfotyrosine (Potential).
FT   MOD_RES    236    236       Sulfotyrosine (Potential).
FT   MOD_RES    240    240       Sulfotyrosine (Potential).
FT   MOD_RES    262    262       Sulfotyrosine (Potential).
FT   MOD_RES    263    263       Sulfotyrosine (Potential).
FT   MOD_RES    338    338       Sulfotyrosine (Potential).
FT   MOD_RES    340    340       Sulfotyrosine (Potential).
FT   MOD_RES    346    346       Sulfotyrosine (Potential).
FT   MOD_RES    347    347       Sulfotyrosine (Potential).
FT   MOD_RES    416    416       Sulfotyrosine (Potential).
FT   MOD_RES    417    417       Sulfotyrosine (Potential).
FT   MOD_RES    420    420       Sulfotyrosine (Potential).
FT   MOD_RES    421    421       Sulfotyrosine (Potential).
FT   MOD_RES    570    570       Hydroxyproline.
FT   MOD_RES    576    576       Hydroxyproline.
FT   MOD_RES    621    621       Hydroxyproline.
FT   MOD_RES    627    627       5-hydroxylysine.
FT   MOD_RES    639    639       Hydroxyproline.
FT   MOD_RES    642    642       5-hydroxylysine.
FT   MOD_RES    648    648       Hydroxyproline.
FT   MOD_RES    654    654       Hydroxyproline.
FT   MOD_RES    657    657       Hydroxyproline.
FT   MOD_RES    675    675       Hydroxyproline.
FT   MOD_RES    678    678       Hydroxyproline.
FT   MOD_RES    680    680       Hydroxyproline.
FT   MOD_RES    686    686       Hydroxyproline.
FT   MOD_RES    690    690       Hydroxyproline.
FT   MOD_RES    696    696       Hydroxyproline.
FT   MOD_RES    705    705       Hydroxyproline.
FT   MOD_RES    708    708       5-hydroxylysine.
FT   MOD_RES    717    717       Hydroxyproline.
FT   MOD_RES    720    720       Hydroxyproline.
FT   MOD_RES    726    726       Hydroxyproline.
FT   MOD_RES    732    732       Hydroxyproline.
FT   MOD_RES    744    744       5-hydroxylysine.
FT   MOD_RES    750    750       Hydroxyproline.
FT   MOD_RES    756    756       Hydroxyproline.
FT   MOD_RES    762    762       Hydroxyproline.
FT   MOD_RES    765    765       Hydroxyproline.
FT   MOD_RES    771    771       Hydroxyproline.
FT   MOD_RES    774    774       5-hydroxylysine.
FT   MOD_RES    780    780       Hydroxyproline.
FT   MOD_RES    789    789       Hydroxyproline.
FT   MOD_RES    795    795       5-hydroxylysine.
FT   MOD_RES    804    804       5-hydroxylysine.
FT   MOD_RES    807    807       5-hydroxylysine.
FT   MOD_RES    810    810       5-hydroxylysine.
FT   MOD_RES    816    816       Hydroxyproline.
FT   MOD_RES    819    819       5-hydroxylysine.
FT   MOD_RES    834    834       Hydroxyproline.
FT   MOD_RES    846    846       5-hydroxylysine.
FT   MOD_RES    861    861       Hydroxyproline.
FT   MOD_RES    864    864       5-hydroxylysine.
FT   MOD_RES    870    870       Hydroxyproline.
FT   MOD_RES    873    873       Hydroxyproline.
FT   MOD_RES    876    876       Hydroxyproline.
FT   MOD_RES    882    882       5-hydroxylysine.
FT   MOD_RES    888    888       Hydroxyproline.
FT   MOD_RES    891    891       Hydroxyproline.
FT   MOD_RES    897    897       5-hydroxylysine.
FT   MOD_RES    903    903       Hydroxyproline.
FT   MOD_RES    906    906       Hydroxyproline.
FT   MOD_RES    930    930       Hydroxyproline.
FT   MOD_RES    945    945       Hydroxyproline.
FT   MOD_RES   1017   1017       Hydroxyproline.
FT   MOD_RES   1020   1020       Hydroxyproline.
FT   MOD_RES   1023   1023       Hydroxyproline.
FT   MOD_RES   1029   1029       Hydroxyproline.
FT   MOD_RES   1221   1221       Hydroxyproline.
```

```
FT   MOD_RES     1224   1224       Hydroxyproline.
FT   MOD_RES     1467   1467       Hydroxyproline.

  Query Match             13.4%;  Score 280.5;  DB 1;  Length 1838;
  Best Local Similarity   30.1%;  Pred. No. 4.8e-08;
  Matches   83;  Conservative    9;  Mismatches   69;  Indels  115;  Gaps   10;

Qy         13 PAAAPRERGSQGC-GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRE 71
              |    | |:|  |  |  | | | |    |    :|| |
Db        857 PLGPPGEKGKLGVPGLPGYPGRQGPKGS----IGFPGFP-------------------- 891

Qy         72 RGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGH 131
               ||    || |||                        |:| |: |    |
Db        892 -GANGEKGGRGTP---------------------GKPGPRGQRGPTG------------ 916

Qy        132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPP---GPNGPP 188
                          |  |   | :      | | | |||          ||||   |||||
Db        917 -------------PRGERGPRGITGKPGPKGNSGGDGP---------AGPPGERGPNGPQ 954

Qy        189 GPPGPPGPQGPPGIPGIPGIPG-------TTVMGPPGPPGP------------------- 222
              || | |||:|||| ||  |:||       |    |  |||||
Db        955 GPTGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGER 1014

Qy        223 --PGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQ 256
                 ||| |||| ||  | | | ||:   |    | |:
Db       1015 GHPGPPGPPGEQGLPGLAGKEGTK-GDPGPAGLPGK 1049


RESULT 15
Q15094_HUMAN
ID   Q15094_HUMAN PRELIMINARY;      PRT;  1838 AA.
AC   Q15094;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Pro-alpha-1 type V collagen.
GN   Name=COL5A1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=92105142; PubMed=1722213;
RA   Greenspan D.S., Cheng W., Hoffman G.G.;
RT   "The pro-alpha 1(V) collagen chain. Complete primary structure,
RT   distribution of expression, and comparison with the pro-alpha 1(XI)
RT   collagen chain.";
RL   J. Biol. Chem. 266:24727-24733(1991).
CC   -!- FUNCTION: May play an important role in fibrillogenesis by
CC       controlling lateral growth of collagen II fibrils (By similarity).
CC   -!- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),
CC       alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
CC       modification of alpha 1(II). Alpha 1(V) can also be found instead
CC       of alpha 3(XI)=1(II) (By similarity).
DR   EMBL; M76729; AAA59993.1; -; mRNA.
DR   PIR; S18802; CGHU1V.
DR   Ensembl; ENSG00000130635; Homo sapiens.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; Laminin_G_TSP_N.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen; Extracellular matrix; Structural protein.
SQ   SEQUENCE   1838 AA;  183612 MW;  5078307F6E00F0BA CRC64;

  Query Match             13.4%;  Score 280.5;  DB 2;  Length 1838;
  Best Local Similarity   30.1%;  Pred. No. 4.8e-08;
  Matches   83;  Conservative    9;  Mismatches   69;  Indels  115;  Gaps   10;

Qy         13 PAAAPRERGSQGC-GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRE 71
              |    | |:|  |  |  | | | |    |    :|| |
Db        857 PLGPPGEKGKLGVPGLPGYPGRQGPKGS----IGFPGFP-------------------- 891

Qy         72 RGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGH 131
               ||    || |||                        |:| |: |    |
Db        892 -GANGEKGGRGTP---------------------GKPGPRGQRGPTG------------ 916

Qy        132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPP---GPNGPP 188
                          |  |   | :      | | | |||          ||||   |||||
Db        917 -------------PRGERGPRGITGKPGPKGNSGGDGP---------AGPPGERGPNGPQ 954

Qy        189 GPPGPPGPQGPPGIPGIPGIPG-------TTVMGPPGPPGP------------------- 222
              || | |||:|||| ||  |:||       |    |  |||||
Db        955 GPTGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGER 1014

Qy        223 --PGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQ 256
                 ||| |||| ||  | | | ||:   |    | |:
Db       1015 GHPGPPGPPGEQGLPGLAGKEGTK-GDPGPAGLPGK 1049


Search completed: January 13, 2006, 16:14:55
Job time : 75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2006, 16:11:42 ; Search time 23 Seconds
(without alignments)
1405.487 Million cell updates/sec

Title: US-09-729-658C-2
Perfect score: 2098
Sequence: 1 MGYPEVERRELLPAAAPRER.....NMSKHTTFFGAIRLGEAPAS 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2098	100.0	391	2	US-09-342-681C-2	Sequence 2, Appli
2	2098	100.0	391	2	US-10-012-605C-6	Sequence 6, Appli
3	2098	100.0	391	2	US-09-949-016-6074	Sequence 6074, Ap
4	1873	89.3	377	2	US-09-342-681C-4	Sequence 4, Appli
5	1403	66.9	260	2	US-09-949-016-11119	Sequence 11119, A
6	294	14.0	357	1	US-07-609-716-66	Sequence 66, Appl
7	294	14.0	357	1	US-08-642-255-33	Sequence 33, Appl
8	294	14.0	357	2	US-08-475-411A-66	Sequence 66, Appl
9	294	14.0	357	2	US-08-478-029A-66	Sequence 66, Appl
10	277	13.2	330	1	US-08-642-255-32	Sequence 32, Appl
11	277	13.2	408	1	US-07-609-716-65	Sequence 65, Appl
12	277	13.2	408	2	US-08-475-411A-65	Sequence 65, Appl
13	277	13.2	408	2	US-08-478-029A-65	Sequence 65, Appl
14	265.5	12.7	1057	2	US-08-931-820-1	Sequence 1, Appli
15	265.5	12.7	1057	2	US-10-153-469A-16	Sequence 16, Appl
16	265.5	12.7	1057	2	US-10-153-469A-20	Sequence 20, Appl
17	265.5	12.7	1057	2	US-10-104-889-16	Sequence 16, Appl
18	265.5	12.7	1057	2	US-10-104-889-20	Sequence 20, Appl
19	265.5	12.7	1064	1	US-08-642-255-62	Sequence 62, Appl
20	265.5	12.7	1107	2	US-10-153-469A-11	Sequence 11, Appl
21	265.5	12.7	1107	2	US-10-104-889-11	Sequence 11, Appl
22	265.5	12.7	1169	2	US-10-153-469A-6	Sequence 6, Appli
23	265.5	12.7	1169	2	US-10-104-889-6	Sequence 6, Appli
24	265.5	12.7	1171	2	US-10-153-469A-8	Sequence 8, Appli
25	265.5	12.7	1171	2	US-10-104-889-8	Sequence 8, Appli
26	265.5	12.7	1341	2	US-08-963-825-18	Sequence 18, Appl
27	265.5	12.7	1341	2	US-09-500-811-18	Sequence 18, Appl
28	265.5	12.7	1341	2	US-09-570-573-18	Sequence 18, Appl
29	265.5	12.7	1341	2	US-09-548-608-18	Sequence 18, Appl
30	265.5	12.7	1388	2	US-10-153-469A-10	Sequence 10, Appl
31	265.5	12.7	1388	2	US-10-104-889-10	Sequence 10, Appl
32	265.5	12.7	1461	2	US-09-585-887-9	Sequence 9, Appli
33	265.5	12.7	1461	2	US-09-289-578-9	Sequence 9, Appli
34	265.5	12.7	1464	2	US-09-331-347C-21	Sequence 21, Appl
35	264	12.6	1218	2	US-09-949-016-7065	Sequence 7065, Ap
36	264	12.6	1690	2	US-09-949-016-5884	Sequence 5884, Ap
37	263.5	12.6	504	2	US-09-219-849-3	Sequence 3, Appli
38	263.5	12.6	561	1	US-08-642-255-52	Sequence 52, Appl
39	263.5	12.6	720	2	US-09-219-849-4	Sequence 4, Appli
40	263.5	12.6	777	1	US-08-642-255-53	Sequence 53, Appl
41	263.5	12.6	1442	1	US-08-316-650-12	Sequence 12, Appl
42	263.5	12.6	1442	4	PCT-US95-02251-12	Sequence 12, Appl
43	263	12.5	1024	2	US-08-931-820-2	Sequence 2, Appli
44	263	12.5	1366	2	US-08-963-825-19	Sequence 19, Appl
45	263	12.5	1366	2	US-09-500-811-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-342-681C-2
; Sequence 2, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
; APPLICANT: Zonana et al.
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
; FILE REFERENCE: 52978
; CURRENT APPLICATION NUMBER: US/09/342,681C
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/092,279
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/112,366
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-681C-2

Query Match 100.0%; Score 2098; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.4e-168;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
```

```
Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 2
US-10-012-605C-6
; Sequence 6, Application US/10012605C
; Patent No. 6692748
; GENERAL INFORMATION:
;  APPLICANT: Haldeman, Betty A.
;  APPLICANT:  Thayer, Edward C.
;  APPLICANT:  Sheppard, Paul O.
;  TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
;  TITLE OF INVENTION:  ZACRP3X2
;  FILE REFERENCE: 00-111
;  CURRENT APPLICATION NUMBER: US/10/012,605C
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: US 60/254,019
;  PRIOR FILING DATE: 2000-12-07
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-605C-6

  Query Match           100.0%;  Score 2098;  DB 2;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-168;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 3
US-09-949-016-6074
; Sequence 6074, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6074
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-6074

  Query Match           100.0%;  Score 2098;  DB 2;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-168;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 4
US-09-342-681C-4
; Sequence 4, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
;  APPLICANT: Zonana et al.
;  TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
;  FILE REFERENCE: 52978
;  CURRENT APPLICATION NUMBER: US/09/342,681C
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER: 60/092,279
;  PRIOR FILING DATE: 1998-07-09
;  PRIOR APPLICATION NUMBER: 60/112,366
;  PRIOR FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 123
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 377
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-342-681C-4

  Query Match            89.3%;  Score 1873;  DB 2;  Length 377;
  Best Local Similarity   90.8%;  Pred. No. 1.2e-149;
  Matches  355;  Conservative    2;  Mismatches   20;  Indels   14;  Gaps    1;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
```

```
                ||||||||||| |||||||||||||||| |||||||||||| ||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
               |||||||||||| |||||| | |||||||||| | |||  ||| |||||| :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                 |  :||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db        121 DPLPPESQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGEL------- 293

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 -------EVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 346

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
               |||||||||||||||||||||||||||||
Db        347 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 377


RESULT 5
US-09-949-016-11119
; Sequence 11119, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11119
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-11119

  Query Match           66.9%;  Score 1403;  DB 2;  Length 260;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-110;
  Matches  260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPP 191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPP 60

Qy        192 GPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVV 251
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVV 120

Qy        252 HLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYY 311
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYY 180

Qy        312 INFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISI 371
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 INFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISI 240

Qy        372 NMSKHTTFFGAIRLGEAPAS 391
               ||||||||||||||||||||
Db        241 NMSKHTTFFGAIRLGEAPAS 260


RESULT 6
US-07-609-716-66
; Sequence 66, Application US/07609716
; Patent No. 5514581
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  118
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/609,716
;      FILING DATE:  06-NOV-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A-55186-3/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-609-716-66

  Query Match           14.0%;  Score 294;  DB 1;  Length 357;
  Best Local Similarity  32.4%;  Pred. No. 1.1e-16;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy         71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              :||      |  |:|| :| : | |   |  |       | | |   |  ||   |
Db         71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGLPGPKG 130

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
              |                          | :  |  :|: ||          ||||| |||
Db        131 D-------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy        189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
              |||| ||| |||| || ||:||
Db        166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy        213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
               | |||||||||| ||||| || |     || :
Db        226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258


RESULT 7
US-08-642-255-33
; Sequence 33, Application US/08642255
; Patent No. 5773249
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  CAPPELLO, Joseph
;    APPLICANT:  FERRARI, Franco A.
;    TITLE OF INVENTION:  High Molecular Weight Collagen-Like
;    TITLE OF INVENTION:  Protein Polymers
;    NUMBER OF SEQUENCES:  135
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/642,255
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ROWLAND, Bertram I.
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  A55556-3/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299 FHT UR
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-642-255-33

  Query Match             14.0%;  Score 294;  DB 1;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 1.1e-16;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy          71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
               :||      |  |:|| :| : | |   | |       | | |   |  ||   |
Db          71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGLPGPKG 130

Qy         129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
               |                          | :   |  :|: ||           ||||| |||
Db         131 D------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy         189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
               |||| ||| |||| || ||:||
Db         166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy         213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
                | |||||||||| ||||| || |    || :
Db         226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258


RESULT 8
US-08-475-411A-66
; Sequence 66, Application US/08475411A
; Patent No. 6140072
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  119
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/475,411A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/609,716
;      FILING DATE:  06-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/269,429
;      FILING DATE:  09-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/114,618
;      FILING DATE:  29-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/927,258
;      FILING DATE:  04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-55186-9/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-475-411A-66

  Query Match             14.0%;  Score 294;  DB 2;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 1.1e-16;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy          71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
               :||      |  |:|| :| : | |   | |       | | |   |  ||   |
Db          71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGLPGPKG 130

Qy         129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
               |                          | :   |  :|: ||           ||||| |||
Db         131 D------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy         189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
               |||| ||| |||| || ||:||
Db         166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy         213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
                | |||||||||| ||||| || |    || :
Db         226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258


RESULT 9
US-08-478-029A-66
; Sequence 66, Application US/08478029A
; Patent No. 6184348
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  119
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
```

```
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,029A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/609,716
;      FILING DATE:  06-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/269,429
;      FILING DATE:  09-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/114,618
;      FILING DATE:  29-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/927,258
;      FILING DATE:  04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-55186-8/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-478-029A-66

  Query Match             14.0%;  Score 294;  DB 2;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 1.1e-16;
  Matches   69;  Conservative    9;  Mismatches   71;  Indels   64;  Gaps    3;

Qy         71 ERGAESRLGGSGTPGTSGTLSSLGGLDPDSP--ITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              :||     |  |:|| :| : | |   |  |     | | |   |  ||   |
Db         71 DRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPLPGPKG 130

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
              |                          | :   |  :|: ||          ||||| |||
Db        131 D------------------------RGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 165

Qy        189 GPPGPPGPQGPPGIPGIPGIPGTT---------------------------------- 212
              |||| ||| |||| || ||:||
Db        166 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPP 225

Qy        213 -VMGPPGPPGPPGPQGPPGLQGPSGAADKAGTR 244
               | |||||||| ||||| || |    || :
Db        226 GPPGAPGPPGPPGPPGPPGLPGPKGDRGDAGPK 258


RESULT 10
US-08-642-255-32
; Sequence 32, Application US/08642255
; Patent No. 5773249
;  GENERAL INFORMATION:
;    APPLICANT:  CAPPELLO, Joseph
;    APPLICANT:  FERRARI, Franco A.
;    TITLE OF INVENTION:  High Molecular Weight Collagen-Like
;    TITLE OF INVENTION:  Protein Polymers
;    NUMBER OF SEQUENCES:  135
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/642,255
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ROWLAND, Bertram I.
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  A55556-3/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299 FHT UR
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  330 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-642-255-32

  Query Match             13.2%;  Score 277;  DB 1;  Length 330;
  Best Local Similarity   39.4%;  Pred. No. 2.7e-15;
  Matches   65;  Conservative    5;  Mismatches   61;  Indels   34;  Gaps    4;

Qy         79 GGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNF 138
              | | || :| : | |   | |     | | |   |  ||
Db         58 GPPGPPGPAGPVGSPGAPGPPGP----PGPPGPPGAPGPPG------------------- 94

Qy        139 FFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGPPGPQG 198
                |  |    :  |       : || ||           | ||| ||||||||||| |
Db         95 --PPGPPGPPGPAGPV-------GSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGPAG 145

Qy        199 PPGIPGIPGIPGTTVMGPPGPPGPPGPPGPQGPPGLQGPSGAADKAGT 243
              | | || || ||    ||||||| ||| |||| || | |   |:
Db        146 PVGSPGAPGPPGPP--GPPGPPGAPGPPGPPGPPGPPGPAGPVGS 188


RESULT 11
US-07-609-716-65
; Sequence 65, Application US/07609716
; Patent No. 5514581
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  118
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/07/609,716
;      FILING DATE:  06-NOV-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A-55186-3/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  65:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  408 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-609-716-65

  Query Match             13.2%;  Score 277;  DB 1;  Length 408;
  Best Local Similarity   39.4%;  Pred. No. 3.5e-15;
  Matches   65;  Conservative    5;  Mismatches   61;  Indels   34;  Gaps    4;

Qy         79 GGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNF 138
              |  | || :| : | |   |  |     | | |   |  ||
Db         58 GPPGPPGPAGPVGSPGAPGPPGP----PGPPGPPGAPGPPG------------------- 94

Qy        139 FFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGPPGPQG 198
                |    |      :  |       : || ||          | ||| ||||||||||| |
Db         95 --PPGPPGPPGPAGPV-------GSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGPAG 145

Qy        199 PPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGT 243
              | | || || ||    ||||||| ||| |||| || | |   |:
Db        146 PVGSPGAPGPPGPP--GPPGPPGAPGPPGPPGPPGPPGPAGPVGS 188


RESULT 12
US-08-475-411A-65
; Sequence 65, Application US/08475411A
; Patent No. 6140072
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  119
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/475,411A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/609,716
;      FILING DATE:  06-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/269,429
;      FILING DATE:  09-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/114,618
;      FILING DATE:  29-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/927,258
;      FILING DATE:  04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-55186-9/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  65:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  408 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-475-411A-65

  Query Match             13.2%;  Score 277;  DB 2;  Length 408;
  Best Local Similarity   39.4%;  Pred. No. 3.5e-15;
  Matches   65;  Conservative    5;  Mismatches   61;  Indels   34;  Gaps    4;

Qy         79 GGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNF 138
              |  | || :| : | |   |  |     | | |   |  ||
Db         58 GPPGPPGPAGPVGSPGAPGPPGP----PGPPGPPGAPGPPG------------------- 94

Qy        139 FFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGPPGPQG 198
                |    |      :  |       : || ||          | ||| ||||||||||| |
Db         95 --PPGPPGPPGPAGPV-------GSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGPAG 145

Qy        199 PPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGT 243
              | | || || ||    ||||||| ||| |||| || | |   |:
Db        146 PVGSPGAPGPPGPP--GPPGPPGAPGPPGPPGPPGPPGPAGPVGS 188


RESULT 13
US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  119
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,029A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/609,716
;      FILING DATE:  06-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/269,429
;      FILING DATE:  09-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/114,618
;      FILING DATE:  29-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/927,258
;      FILING DATE:  04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-55186-8/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  65:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  408 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-478-029A-65

  Query Match           13.2%;  Score 277;  DB 2;  Length 408;
  Best Local Similarity   39.4%;  Pred. No. 3.5e-15;
  Matches   65;  Conservative    5;  Mismatches   61;  Indels   34;  Gaps    4;

Qy         79 GGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGHQMALLNF 138
              |  | || :| : | |   |  |    | | |  |  ||
Db         58 GPPGPPGPAGPVGSPGAPGPPGP----PGPPGPPGAPGPPG------------------- 94

Qy        139 FFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGPPGPQG 198
                |   |     :  |        : || ||          | ||| ||||||||||| |
Db         95 --PPGPPGPPGPAGPV-------GSPGAPGPPGPPGPPGPAGPPGPPGPPGPPGPAG 145

Qy        199 PPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADKAGT 243
              | | || || ||    |||||| ||| |||| || | |   |:
Db        146 PVGSPGAPGPPGPP--GPPGPPGAPGPPGPPGPPGPPGPAGPVGS 188


RESULT 14
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1057 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type I
US-08-931-820-1

  Query Match           12.7%;  Score 265.5;  DB 2;  Length 1057;
  Best Local Similarity   29.7%;  Pred. No. 1.1e-13;
  Matches   78;  Conservative   16;  Mismatches   80;  Indels   89;  Gaps    9;

Qy         13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
              || || ::|  |       | |||  ||    | :     || |
Db        613 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------ 654

Qy         62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
                  ||: : |  | || :|     |   |  |     | | |      ||
Db        655 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 700

Qy        122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                  | :: : |  |          : ||
Db        701 ----------------------------------KGARGSAGPPGATGFPGAAGRVGP 724

Qy        182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
              |||:|  ||||||||      :|| |  |  | ||   :|||||||| | :| ||  ||:
Db        725 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 782

Qy        236 GAADKAGTR--ENQPAVVHLQGQ 256
              ||    | :    |  || | ||
Db        783 GAPGTPGPQGIAGQRGVVGLPGQ 805


RESULT 15
US-10-153-469A-16
; Sequence 16, Application US/10153469A
; Patent No. 6927287
;  GENERAL INFORMATION:
;       APPLICANT: GRUSKIN, ELLIOT A.
;                  BUECHTER, DOUGLAS
;                  BROKAW, JANE
;                  ZHANG, GUANGHUI
;                  PAOLELLA, DAVID
;       TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
;       NUMBER OF SEQUENCES: 50
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: DILWORTH & BARRESE
;            STREET: 333 EARLE OVINGTON BOULEVARD
;            CITY: UNIONDALE
;            STATE: NY
;            COUNTRY: U.S.A.
;            ZIP: 11553
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/153,469A
;            FILING DATE: 22-May-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/169,768
;            FILING DATE: 09-OCT-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: STEEN, JEFFREY S
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (516) 228-8484
;            TELEFAX: (516) 228-8516
;  INFORMATION FOR SEQ ID NO: 16:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1057 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: unknown
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-153-469A-16

  Query Match           12.7%;  Score 265.5;  DB 2;  Length 1057;
  Best Local Similarity   29.7%;  Pred. No. 1.1e-13;
  Matches   78;  Conservative   16;  Mismatches   80;  Indels   89;  Gaps    9;

Qy         13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
              || || ::|  |       | |||  ||    | :     || |
Db        613 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------ 654

Qy         62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
```

```
                    ||: : |  | || :|     |   |  |     | | |      ||
Db        655 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 700

Qy        122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                | :: : |  |         : ||
Db        701 -----------------------------------KGARGSAGPPGATGFPGAAGRVGP 724

Qy        182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
              |||:|  ||||||||      :|| |  |  | ||   :|||||||| | :| ||  ||:
Db        725 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 782

Qy        236 GAADKAGTR--ENQPAVVHLQGQ 256
              ||    | :    |  || | ||
Db        783 GAPGTPGPQGIAGQRGVVGLPGQ 805
```

Search completed: January 13, 2006, 16:15:56
Job time : 24 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2006, 16:15:03 ; Search time 115 Seconds
(without alignments)
1420.620 Million cell updates/sec

Title: US-09-729-658C-2
Perfect score: 2098
Sequence: 1 MGYPEVERRELLPAAAPRER.....NMSKHTTFFGAIRLGEAPAS 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2098	100.0	391	3	US-09-813-329-8	Sequence 8, Appli
2	2098	100.0	391	3	US-09-729-658B-2	Sequence 2, Appli
3	2098	100.0	391	4	US-10-218-547-42	Sequence 42, Appl
4	2098	100.0	391	4	US-10-012-605C-6	Sequence 6, Appli
5	2098	100.0	391	4	US-10-310-793-46	Sequence 46, Appl
6	2098	100.0	391	4	US-10-202-062-42	Sequence 42, Appl
7	2098	100.0	391	4	US-10-719-205-6	Sequence 6, Appli
8	2098	100.0	391	6	US-11-028-780-42	Sequence 42, Appl
9	2098	100.0	391	6	US-11-142-736-8	Sequence 8, Appli
10	1965	93.7	391	3	US-09-813-329-9	Sequence 9, Appli
11	1965	93.7	391	6	US-11-142-736-9	Sequence 9, Appli
12	1873	89.3	377	3	US-09-729-658B-4	Sequence 4, Appli
13	779	37.1	412	5	US-10-503-999-6	Sequence 6, Appli
14	759	36.2	410	5	US-10-503-999-8	Sequence 8, Appli
15	542	25.8	246	4	US-10-185-425-9	Sequence 9, Appli
16	542	25.8	246	5	US-10-871-776-9	Sequence 9, Appli
17	286	13.6	742	5	US-10-820-155-10	Sequence 10, Appl
18	282	13.4	742	4	US-10-203-860-4	Sequence 4, Appli
19	282	13.4	742	5	US-10-820-155-2	Sequence 2, Appli
20	282	13.4	742	5	US-10-820-155-6	Sequence 6, Appli
21	280.5	13.4	1838	5	US-10-852-335A-162	Sequence 162, App
22	280.5	13.4	1838	5	US-10-287-436A-464	Sequence 464, App
23	280.5	13.4	1838	5	US-10-287-436A-1161	Sequence 1161, Ap
24	276	13.2	622	5	US-10-820-155-9	Sequence 9, Appli
25	275.5	13.1	180	5	US-10-797-606-39	Sequence 39, Appl
26	275	13.1	1466	4	US-10-402-089-4	Sequence 4, Appli
27	275	13.1	1466	4	US-10-402-089-6	Sequence 6, Appli
28	275	13.1	1466	4	US-10-402-072A-4	Sequence 4, Appli
29	275	13.1	1466	4	US-10-402-072A-6	Sequence 6, Appli
30	273.5	13.0	1014	5	US-10-901-816A-10	Sequence 10, Appl
31	273	13.0	742	4	US-10-203-860-2	Sequence 2, Appli
32	273	13.0	742	4	US-10-341-434-85	Sequence 85, Appl
33	273	13.0	742	5	US-10-820-155-8	Sequence 8, Appli
34	273	13.0	742	5	US-10-820-155-17	Sequence 17, Appl
35	273	13.0	1497	4	US-10-157-031-128	Sequence 128, App
36	268.5	12.8	342	4	US-10-223-085-28	Sequence 28, Appl
37	268.5	12.8	342	4	US-10-223-084-28	Sequence 28, Appl
38	268.5	12.8	342	4	US-10-223-088-28	Sequence 28, Appl
39	268.5	12.8	342	4	US-10-223-090-28	Sequence 28, Appl
40	268.5	12.8	342	4	US-10-223-087-28	Sequence 28, Appl
41	268.5	12.8	342	4	US-10-223-083-28	Sequence 28, Appl
42	268.5	12.8	342	4	US-10-223-089-28	Sequence 28, Appl
43	268.5	12.8	342	4	US-10-223-081-28	Sequence 28, Appl
44	268.5	12.8	342	4	US-10-223-082-28	Sequence 28, Appl
45	268.5	12.8	342	4	US-10-305-654-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-813-329-8
; Sequence 8, Application US/09813329
; Patent No. US20020012968A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Suibb Company
; TITLE OF INVENTION: No. US20020012968A1el Drosophila Tumor Necrosis Factor Class Mole
; TITLE OF INVENTION: Variants Thereof
; FILE REFERENCE: D0016.np
; CURRENT APPLICATION NUMBER: US/09/813,329
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,816
; PRIOR FILING DATE: 2000-03-21
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-813-329-8

Query Match 100.0%; Score 2098; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.6e-137;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
```

```
                                  ||||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 2
US-09-729-658B-2
; Sequence 2, Application US/09729658B
; Publication No. US20030023991A1
; GENERAL INFORMATION:
;  APPLICANT: Zonana et al.
;  TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
;  FILE REFERENCE: 55924
;  CURRENT APPLICATION NUMBER: US/09/729,658B
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/342,681
;  PRIOR FILING DATE: 1999-06-29
;  PRIOR APPLICATION NUMBER: 60/092,279
;  PRIOR FILING DATE: 1998-07-09
;  PRIOR APPLICATION NUMBER: 60/112,366
;  PRIOR FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 122
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-729-658B-2

  Query Match           100.0%;  Score 2098;  DB 3;  Length 391;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              ||||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 3
US-10-218-547-42
; Sequence 42, Application US/10218547
; Publication No. US20030100074A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Methods And Compositions For Treating Metabolic Bone Diseases Rel
;  TITLE OF INVENTION:  Human Endokine Alpha
;  FILE REFERENCE: PF561
;  CURRENT APPLICATION NUMBER: US/10/218,547
;  CURRENT FILING DATE:  2002-08-15
;  PRIOR APPLICATION NUMBER: 60/312,542
;  PRIOR FILING DATE: 2001-08-16
;  PRIOR APPLICATION NUMBER: 60/330,761
;  PRIOR FILING DATE: 2001-10-30
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: human
US-10-218-547-42

  Query Match           100.0%;  Score 2098;  DB 4;  Length 391;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              ||||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 4
US-10-012-605C-6
; Sequence 6, Application US/10012605C
; Publication No. US20030104602A1
; GENERAL INFORMATION:
;  APPLICANT: Haldeman, Betty A.
;  APPLICANT:  Thayer, Edward C.
;  APPLICANT:  Sheppard, Paul O.
;  TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
;  TITLE OF INVENTION:  ZACRP3X2
;  FILE REFERENCE: 00-111
;  CURRENT APPLICATION NUMBER: US/10/012,605C
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: US 60/254,019
;  PRIOR FILING DATE: 2000-12-07
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-605C-6

  Query Match           100.0%;  Score 2098;  DB 4;  Length 391;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 5
US-10-310-793-46
; Sequence 46, Application US/10310793
; Publication No. US20030198640A1
; GENERAL INFORMATION:
;  APPLICANT: Yu, Guo-Liang
;  APPLICANT:  Ni, Jian
;  APPLICANT:  Rosen, Craig A
;  APPLICANT:  Zhang, Jun
;  APPLICANT:  Wei, Ping
;  TITLE OF INVENTION: Methods And Compositions For Treating Inflammatory Bowel Diseases
;  TITLE OF INVENTION:  Relating To Human Tumor Necrosis Factor-Gamma Beta
;  FILE REFERENCE: PF573
;  CURRENT APPLICATION NUMBER: US/10/310,793
;  CURRENT FILING DATE:  2002-12-06
;  PRIOR APPLICATION NUMBER: 60/336,695
;  PRIOR FILING DATE: 2001-12-07
;  PRIOR APPLICATION NUMBER: 10/226,294
;  PRIOR FILING DATE: 2002-08-23
;  PRIOR APPLICATION NUMBER: 60/314,381
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: 09/899,059
;  PRIOR FILING DATE: 2001-07-06
;  PRIOR APPLICATION NUMBER: 60/278,449
;  PRIOR FILING DATE: 2001-03-26
;  PRIOR APPLICATION NUMBER: 60/216,879
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 09/559,290
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/180,908
;  PRIOR FILING DATE: 2000-02-08
;  PRIOR APPLICATION NUMBER: 60/134,067
;  PRIOR FILING DATE: 1999-05-13
;  PRIOR APPLICATION NUMBER: 60/132,227
;  PRIOR FILING DATE: 1999-05-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 71
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: human
US-10-310-793-46

  Query Match           100.0%;  Score 2098;  DB 4;  Length 391;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 6
US-10-202-062-42
; Sequence 42, Application US/10202062
; Publication No. US20040038349A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.,
;  TITLE OF INVENTION: Heteromultimeric TNF Ligand Family members
;  FILE REFERENCE: PF559
;  CURRENT APPLICATION NUMBER: US/10/202,062
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/307,838
;  PRIOR FILING DATE: 2001-07-27
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: human
US-10-202-062-42

  Query Match           100.0%;  Score 2098;  DB 4;  Length 391;
  Best Local Similarity  100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 7
US-10-719-205-6
; Sequence 6, Application US/10719205
; Publication No. US20040086971A1
; GENERAL INFORMATION:
;  APPLICANT: Haldeman, Betty A.
;  APPLICANT:  Thayer, Edward C.
;  APPLICANT:  Sheppard, Paul O.
;  TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
;  TITLE OF INVENTION:  ZACRP3X2
;  FILE REFERENCE: 00-111
;  CURRENT APPLICATION NUMBER: US/10/719,205
;  CURRENT FILING DATE:  2003-11-21
;  PRIOR APPLICATION NUMBER: US/10/012,605C
;  PRIOR FILING DATE: 2002-08-14
;  PRIOR APPLICATION NUMBER: US 60/254,019
;  PRIOR FILING DATE: 2000-12-07
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-719-205-6

  Query Match             100.0%;  Score 2098;  DB 4;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 8
US-11-028-780-42
; Sequence 42, Application US/11028780
; Publication No. US20050163747A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.,
;  TITLE OF INVENTION: Heteromultimeric TNF Ligand Family members
;  FILE REFERENCE: PF559C1
;  CURRENT APPLICATION NUMBER: US/11/028,780
;  CURRENT FILING DATE:  2005-01-05
;  PRIOR APPLICATION NUMBER: 10/202,062
;  PRIOR FILING DATE: 2002-07-25
;  PRIOR APPLICATION NUMBER: 60/307,838
;  PRIOR FILING DATE: 2001-07-27
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: human
US-11-028-780-42

  Query Match             100.0%;  Score 2098;  DB 6;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 9
US-11-142-736-8
; Sequence 8, Application US/11142736
; Publication No. US20050227283A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Suibb Company
;  TITLE OF INVENTION: Novel Drosophila Tumor Necrosis Factor Class Molecule ("DmTNF") a
;  TITLE OF INVENTION:  Variants Thereof
;  FILE REFERENCE: D0016 DIV1
;  CURRENT APPLICATION NUMBER: US/11/142,736
;  CURRENT FILING DATE:  2005-06-01
;  PRIOR APPLICATION NUMBER: 60/190,816
;  PRIOR FILING DATE: 2000-03-21
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-11-142-736-8

  Query Match             100.0%;  Score 2098;  DB 6;  Length 391;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.6e-137;
  Matches  391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 10
US-09-813-329-9
; Sequence 9, Application US/09813329
; Patent No. US20020012968A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Suibb Company
;  TITLE OF INVENTION: No. US20020012968A1el Drosophila Tumor Necrosis Factor Class Mole
;  TITLE OF INVENTION:  Variants Thereof
;  FILE REFERENCE: D0016.np
;  CURRENT APPLICATION NUMBER: US/09/813,329
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: 60/190,816
;  PRIOR FILING DATE: 2000-03-21
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-813-329-9

  Query Match           93.7%;  Score 1965;  DB 3;  Length 391;
  Best Local Similarity   94.6%;  Pred. No. 4.3e-128;
  Matches  370;  Conservative    1;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| |||||||||||||| |||||||||| |||||||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||| |||||| | ||||||||| | |||  ||| |||||| :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                |  |||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db        121 DPLPPDSQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 11
US-11-142-736-9
; Sequence 9, Application US/11142736
; Publication No. US20050227283A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Suibb Company
;  TITLE OF INVENTION: Novel Drosophila Tumor Necrosis Factor Class Molecule ("DmTNF") a
;  TITLE OF INVENTION:  Variants Thereof
;  FILE REFERENCE: D0016 DIV1
;  CURRENT APPLICATION NUMBER: US/11/142,736
;  CURRENT FILING DATE:  2005-06-01
;  PRIOR APPLICATION NUMBER: 60/190,816
;  PRIOR FILING DATE: 2000-03-21
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-142-736-9

  Query Match           93.7%;  Score 1965;  DB 6;  Length 391;
  Best Local Similarity   94.6%;  Pred. No. 4.3e-128;
  Matches  370;  Conservative    1;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| |||||||||||||| |||||||||| |||||||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||| |||||| | ||||||||| | |||  ||| |||||| :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                |  |||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db        121 DPLPPDSQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391


RESULT 12
US-09-729-658B-4
; Sequence 4, Application US/09729658B
; Publication No. US20030023991A1
; GENERAL INFORMATION:
;  APPLICANT: Zonana et al.
```

```
;  TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
;  FILE REFERENCE: 55924
;  CURRENT APPLICATION NUMBER: US/09/729,658B
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/342,681
;  PRIOR FILING DATE: 1999-06-29
;  PRIOR APPLICATION NUMBER: 60/092,279
;  PRIOR FILING DATE: 1998-07-09
;  PRIOR APPLICATION NUMBER: 60/112,366
;  PRIOR FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 122
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 377
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-729-658B-4

  Query Match             89.3%;  Score 1873;  DB 3;  Length 377;
  Best Local Similarity   90.8%;  Pred. No. 1e-121;
  Matches  355;  Conservative    2;  Mismatches   20;  Indels   14;  Gaps    1;

Qy          1 MGYPEVERRELLPAAAPRERGSQGCGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCC 60
              |||||||||| |||||||||||||||| |||||||||||| ||||||||||||||||||||
Db          1 MGYPEVERREPLPAAAPRERGSQGCGCRGAPARAGEGNSCRLFLGFFGLSLALHLLTLCC 60

Qy         61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
              |||||||||||||| |||||| | |||||||||| | |||  ||| |||||| :||||||||
Db         61 YLELRSELRRERGTESRLGGPGAPGTSGTLSSPGSLDPVGPITRHLGQPSFQQQPLEPGE 120

Qy        121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                |  :||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Db        121 DPLPPESQDRHQMALLNFFFPDEKAYSEEESRRVRRNKRSKSGEGADGPVKNKKKGKKAG 180

Qy        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPSGAADK 240

Qy        241 AGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGELEVLVDGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDWSRITMNPKVFKLHPRSGEL------- 293

Qy        301 YFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 360
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 -------EVYYINFTDFASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKI 346

Qy        361 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||
Db        347 AVKMVHADISINMSKHTTFFGAIRLGEAPAS 377


RESULT 13
US-10-503-999-6
; Sequence 6, Application US/10503999
; Publication No. US20050152872A1
; GENERAL INFORMATION:
;  APPLICANT: Gaide, Olivier
;  APPLICANT:  Schneider, Pascal
;  APPLICANT:  Tschopp, Jurg
;  TITLE OF INVENTION: Fusion Constructs Containing Active Sections Of TNF Ligands
;  FILE REFERENCE: 11436*13
;  CURRENT APPLICATION NUMBER: US/10/503,999
;  CURRENT FILING DATE:  2004-08-09
;  PRIOR APPLICATION NUMBER: PCT/EP02/09354
;  PRIOR FILING DATE: 2002-08-21
;  PRIOR APPLICATION NUMBER: DE 102 05 368.5
;  PRIOR FILING DATE: 2002-02-10
;  PRIOR APPLICATION NUMBER: DE 102 05 583.1
;  PRIOR FILING DATE: 2002-02-11
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
;   LENGTH: 412
;   TYPE: PRT
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-10-503-999-6

  Query Match             37.1%;  Score 779;  DB 5;  Length 412;
  Best Local Similarity   87.6%;  Pred. No. 7.9e-46;
  Matches  155;  Conservative    1;  Mismatches   15;  Indels    6;  Gaps    2;

Qy        215 GPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 274
              |   |   | ||  |    | |:       ||||||||||||||||||||||||||||||
Db        242 GKRSPQPQPKPQPKP---EPEGSLQ---VDENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 295

Qy        275 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTR 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTR 355

Qy        335 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 412


RESULT 14
US-10-503-999-8
; Sequence 8, Application US/10503999
; Publication No. US20050152872A1
; GENERAL INFORMATION:
;  APPLICANT: Gaide, Olivier
;  APPLICANT:  Schneider, Pascal
;  APPLICANT:  Tschopp, Jurg
;  TITLE OF INVENTION: Fusion Constructs Containing Active Sections Of TNF Ligands
;  FILE REFERENCE: 11436*13
;  CURRENT APPLICATION NUMBER: US/10/503,999
;  CURRENT FILING DATE:  2004-08-09
;  PRIOR APPLICATION NUMBER: PCT/EP02/09354
;  PRIOR FILING DATE: 2002-08-21
;  PRIOR APPLICATION NUMBER: DE 102 05 368.5
;  PRIOR FILING DATE: 2002-02-10
;  PRIOR APPLICATION NUMBER: DE 102 05 583.1
;  PRIOR FILING DATE: 2002-02-11
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 410
;   TYPE: PRT
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: Synthetic Construct
US-10-503-999-8

  Query Match             36.2%;  Score 759;  DB 5;  Length 410;
  Best Local Similarity   86.4%;  Pred. No. 1.9e-44;
  Matches  153;  Conservative    1;  Mismatches   15;  Indels    8;  Gaps    3;

Qy        215 GPPGPPGPPGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 274
              |   |   | ||  |    | |:       ||||||||||||||||||||||||||||||
Db        242 GKRSPQPQPKPQPKP---EPEGSLQ---VDENQPAVVHLQGQGSAIQVKNDLSGGVLNDW 295

Qy        275 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQVEVYYINFTDFASYEVVVDEKPFLQCTR 334
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        296 SRITMNPKVFKLHPRSGELEVLVDGTYFIYSQ--VYYINFTDFASYEVVVDEKPFLQCTR 353

Qy        335 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 SIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHTTFFGAIRLGEAPAS 410


RESULT 15
US-10-185-425-9
; Sequence 9, Application US/10185425
; Publication No. US20030053984A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Apotech Research and Development Ltd.
;  APPLICANT:  Tschopp, Jurg
;  APPLICANT:  Schneider, Pascal
;  APPLICANT:  Holler, Nils
;  TITLE OF INVENTION: A Bimer or an Oligomer of a Dimer, Trimer, Quadromer, or Pentamer
;  TITLE OF INVENTION:  Recombinant Fusion Proteins
;  FILE REFERENCE: 1708.001US1
;  CURRENT APPLICATION NUMBER: US/10/185,425
;  CURRENT FILING DATE:  2002-06-28
;  PRIOR APPLICATION NUMBER: PCT/EP00/13032
;  PRIOR FILING DATE: 2000-12-20
;  PRIOR APPLICATION NUMBER: DE 199 63 859.4
;  PRIOR FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: EDA-Fas ligand fusion protein
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (1)...(8)
;   OTHER INFORMATION: Flag
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (9)...(16)
;   OTHER INFORMATION: Linker
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (17)...(99)
;   OTHER INFORMATION: Collagen domain aa 160-242
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (100)...(104)
;   OTHER INFORMATION: Linker
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (105)...(246)
;   OTHER INFORMATION: hFas ligand extracellular domain aa 139-281
US-10-185-425-9

  Query Match            25.8%;  Score 542;  DB 4;  Length 246;
  Best Local Similarity  47.6%;  Pred. No. 1.2e-29;
  Matches  118;  Conservative   28;  Mismatches   64;  Indels   38;  Gaps    7;

Qy        159 RSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPG 218
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 QSKSNEGADGPVKNKKKGKKAGPPGPNGPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPPG 75

Qy        219 PPGPPGPQGPPGLQGPSGAADKAGTRENQP----AVVHLQGQGSAIQVK---NDLSGGVL 271
              ||||||||||||||||||||||    ::      | || |: ::  :     |  | ||
Db         76 PPGPPGPQGPPGLQGPSGAADKAGMHVDEKKELRKVAHLTGKSNSRSMPLEWEDTYGIVL 135

Qy        272 NDWSRITMNPKVFKLHPRSGELEVLVDGTYFIYSQV-------------EVYYINFTDF 317
                    :  :  : | | :   | ||:||:|                :||  | : :
Db        136 -----------LSGVKYKKGGLVINETGLYFVYSKVYFRGQSCNNLPLSHKVYMRN-SKY 183

Qy        318 ASYEVVVDEKPFLQCTRSIETGKTNYNTCYTAGVCLLKARQKIAVKMVHADISINMSKHT 377
                  |::: |    ||    ||:    : |   | | :   : | :    : :|  :
Db        184 PQDLVMMEGKMMSYCT----TGQMWARSSYLGAVFNLTSADHLYVNVSELSL-VNFEESQ 238

Qy        378 TFFGAIRL 385
              ||||  :|
Db        239 TFFGLYKL 246


Search completed: January 13, 2006, 16:25:51
Job time : 116 secs
```

```
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 912
;   LENGTH: 1806
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-912

  Query Match             12.1%;  Score 254;  DB 6;  Length 1806;
  Best Local Similarity   32.0%;  Pred. No. 6.2e-12;
  Matches   93;  Conservative   15;  Mismatches   87;  Indels   96;  Gaps   17;

Qy         13 PAAAPRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |   |  || :|     |  |:    ||      | || |  : |          | |:
Db        755 PIGYPGPRGVKGADGVRGLKGSKGEKGEDG----FPGFKG-DMGLK--------GDRGEV 801

Qy         69 -----RRERGAE---SRLGGSGTPGTSGTLSSLGGLD-PDSPITSHLGQPSPKQQPLEPG 119
                   | | | |    | | :| || ||     | |  |  |   : |:  ||     ||
Db        802 GQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLP--GYPGRQGPKGSTGFPG 859

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVR-----RNKRS----KSNEGADGPV 170
                                 ||      |: :| |      | :|     : : || ||
Db        860 --------------------FPGAN--GEKGARGVAGKPGPRGQRGPTGPRGSRGARGPT 897

Qy        171 -KNKKKGKKA-----GPPGPNGPPGPPGP---PGPQGPPGIPGIPGIPG-------TTVM 214
               |   ||         ||||  || || ||    |||:|||| ||  |:||       |
Db        898 GKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQ 957

Qy        215 GPPGPPGP--------------------PGPQGPPGLQGPSGAADKAGTR 244
              |  |||||                    ||| |||| ||  ||| | | :
Db        958 GKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGAAGKEGAK 1008


RESULT 14
US-10-995-561-915
; Sequence 915, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 915
;   LENGTH: 1806
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-915

  Query Match             12.1%;  Score 254;  DB 6;  Length 1806;
  Best Local Similarity   32.0%;  Pred. No. 6.2e-12;
  Matches   93;  Conservative   15;  Mismatches   87;  Indels   96;  Gaps   17;

Qy         13 PAAAPRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |   |  || :|     |  |:    ||      | || |  : |          | |:
Db        755 PIGYPGPRGVKGADGVRGLKGSKGEKGEDG----FPGFKG-DMGLK--------GDRGEV 801

Qy         69 -----RRERGAE---SRLGGSGTPGTSGTLSSLGGLD-PDSPITSHLGQPSPKQQPLEPG 119
                   | | | |    | | :| || ||     | |  |  |   : |:  ||     ||
Db        802 GQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLP--GYPGRQGPKGSTGFPG 859

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVR-----RNKRS----KSNEGADGPV 170
                                 ||      |: :| |      | :|     : : || ||
Db        860 --------------------FPGAN--GEKGARGVAGKPGPRGQRGPTGPRGSRGARGPT 897

Qy        171 -KNKKKGKKA-----GPPGPNGPPGPPGP---PGPQGPPGIPGIPGIPG-------TTVM 214
               |   ||         ||||  || || ||    |||:|||| ||  |:||       |
Db        898 GKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQ 957

Qy        215 GPPGPPGP--------------------PGPQGPPGLQGPSGAADKAGTR 244
              |  |||||                    ||| |||| ||  ||| | | :
Db        958 GKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGAAGKEGAK 1008


Qy        171 -KNKKKGKKA-----GPPGPNGPPGPPGP---PGPQGPPGIPGIPGIPG-------TTVM 214
               |   ||         ||||  || || ||    |||:|||| ||  |:||       |
Db        898 GKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQ 957

Qy        215 GPPGPPGP--------------------PGPQGPPGLQGPSGAADKAGTR 244
              |  |||||                    ||| |||| ||  ||| | | :
Db        958 GKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGAAGKEGAK 1008


RESULT 15
US-10-995-561-910
; Sequence 910, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 910
;   LENGTH: 1818
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-910

  Query Match             12.1%;  Score 254;  DB 6;  Length 1818;
  Best Local Similarity   32.0%;  Pred. No. 6.2e-12;
  Matches   93;  Conservative   15;  Mismatches   87;  Indels   96;  Gaps   17;

Qy         13 PAAAPRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |   |  || :|     |  |:    ||      | || |  : |          | |:
Db        767 PIGYPGPRGVKGADGVRGLKGSKGEKGEDG----FPGFKG-DMGLK--------GDRGEV 813

Qy         69 -----RRERGAE---SRLGGSGTPGTSGTLSSLGGLD-PDSPITSHLGQPSPKQQPLEPG 119
                   | | | |    | | :| || ||     | |  |  |   : |:  ||     ||
Db        814 GQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLP--GYPGRQGPKGSTGFPG 871

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVR-----RNKRS----KSNEGADGPV 170
                                 ||      |: :| |      | :|     : : || ||
Db        872 --------------------FPGAN--GEKGARGVAGKPGPRGQRGPTGPRGSRGARGPT 909

Qy        171 -KNKKKGKKA-----GPPGPNGPPGPPGP---PGPQGPPGIPGIPGIPG-------TTVM 214
               |   ||         ||||  || || ||    |||:|||| ||  |:||       |
Db        910 GKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQ 969

Qy        215 GPPGPPGP--------------------PGPQGPPGLQGPSGAADKAGTR 244
              |  |||||                    ||| |||| ||  ||| | | :
Db        970 GKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGAAGKEGAK 1020


Search completed: January 13, 2006, 16:26:34
Job time : 32 secs
```

```
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/11/186,284
;  CURRENT FILING DATE:  2005-07-21
;  PRIOR APPLICATION NUMBER: US/10/301,822
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 33
;   LENGTH: 1466
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-11-186-284-33

  Query Match             12.2%;  Score 256;  DB 7;  Length 1466;
  Best Local Similarity   28.8%;  Pred. No. 3.4e-12;
  Matches   86;  Conservative   12;  Mismatches   97;  Indels  104;  Gaps   10;

Qy         13 PAAAPRERGSQG----CGCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |  :| |||  |     |  ||| : ||
Db        793 PRGSPGERGETGPPGPAGFPGAPGQNGEPGG---------------------------- 823

Qy         69 RRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              : ||||     |  | ||| :|      |   |  |      |    :  |  || |
Db        824 KGERGAPGEKGEGGPPGVAGPPGGSGPAGPPGP----QGVKGERGSPGGPGAAG------ 873

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
                          ||          :| :     |  | |  ||  :   |     || |  | |
Db        874 -----------FPG--------ARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAP 914

Qy        189 GPP---------------GPPGPQGPPGIPGIPGIPGTT----VMGPPGPPGP---PGPQ 226
              | |               | || ||||| ||  || | |    : |||| |||   ||||
Db        915 GSPGVSGPKGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQ 974

Qy        227 ------------------GPPGLQGPSGAADKAGT--RENQPAVVHLQGQGSAIQVKND 265
                               |||| ||  | |  ||   |:  |     | |: :   | |
Db        975 GVKGESGKPGANGLSGERGPPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGD 1033


RESULT 11
US-10-995-561-911
; Sequence 911, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 911
;   LENGTH: 1767
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-911

  Query Match             12.1%;  Score 254;  DB 6;  Length 1767;
  Best Local Similarity   32.0%;  Pred. No. 6e-12;
  Matches   93;  Conservative   15;  Mismatches   87;  Indels   96;  Gaps   17;

Qy         13 PAAAPRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |  :| |||  |     |  ||| : ||
Db        793 PRGSPGERGETGPPGPAGFPGAPGQNGEPGG---------------------------- 823

Qy         69 RRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQ 128
              : ||||     |  | ||| :|      |   |  |      |    :  |  || |
Db        824 KGERGAPGEKGEGGPPGVAGPPGGSGPAGPPGP----QGVKGERGSPGGPGAAG------ 873

Qy        129 DGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPPGPNGPP 188
                          ||          :| :     |  | |  ||  :   |     || |  | |
Db        874 -----------FPG--------ARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAP 914

Qy        189 GPP---------------GPPGPQGPPGIPGIPGIPGTT----VMGPPGPPGP---PGPQ 226
              | |               | || ||||| ||  || | |    : |||| |||   ||||
Db        915 GSPGVSGPKGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQ 974

Qy        227 ------------------GPPGLQGPSGAADKAGT--RENQPAVVHLQGQGSAIQVKND 265
                               |||| ||  | |  ||   |:  |     | |: :   | |
Db        975 GVKGESGKPGANGLSGERGPPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGD 1033

Qy         13 PAAAPRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |  :| |||  |     |  ||| : ||
Db        716 PIGYPGPRGVKGADGVRGLKGSKGEKGEDG----FPGFKG-DMGLK--------GDRGEV 762

Qy         69 -----RRERGAE---SRLGGSGTPGTSGTLSSLGGLD-PDSPITSHLGQPSPKQQPLEPG 119
                    | | | |     | | :| || ||      |  |  |  |   : |:  ||      ||
Db        763 GQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLP--GYPGRQGPKGSTGFPG 820

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVR-----RNKRS----KSNEGADGPV 170
                                 ||       |: :| |       | :|     : : || ||
Db        821 -------------------FPGAN--GEKGARGVAGKPGPRGQRGPTGPRGSRGARGPT 858

Qy        171 -KNKKKGKKA-----GPPGPNGPPGPPGP---PGPQGPPGIPGIPGIPG-------TTVM 214
                |   ||         ||||  || || ||   |||:|||| ||  |:||       |
Db        859 GKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQ 918

Qy        215 GPPGPPGP--------------------PGPQGPPGLQGPSGAADKAGTR 244
              |  |||||                    ||| |||| ||  ||| | | :
Db        919 GKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGAAGKEGAK 969


RESULT 12
US-10-995-561-914
; Sequence 914, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 914
;   LENGTH: 1767
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-914

  Query Match             12.1%;  Score 254;  DB 6;  Length 1767;
  Best Local Similarity   32.0%;  Pred. No. 6e-12;
  Matches   93;  Conservative   15;  Mismatches   87;  Indels   96;  Gaps   17;

Qy         13 PAAAPRERGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSEL 68
              |  :| |||  |     |  ||| : ||
Db        716 PIGYPGPRGVKGADGVRGLKGSKGEKGEDG----FPGFKG-DMGLK--------GDRGEV 762

Qy         69 -----RRERGAE---SRLGGSGTPGTSGTLSSLGGLD-PDSPITSHLGQPSPKQQPLEPG 119
                    | | | |     | | :| || ||      |  |  |  |   : |:  ||      ||
Db        763 GQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLP--GYPGRQGPKGSTGFPG 820

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVR-----RNKRS----KSNEGADGPV 170
                                 ||       |: :| |       | :|     : : || ||
Db        821 -------------------FPGAN--GEKGARGVAGKPGPRGQRGPTGPRGSRGARGPT 858

Qy        171 -KNKKKGKKA-----GPPGPNGPPGPPGP---PGPQGPPGIPGIPGIPG-------TTVM 214
                |   ||         ||||  || || ||   |||:|||| ||  |:||       |
Db        859 GKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQ 918

Qy        215 GPPGPPGP--------------------PGPQGPPGLQGPSGAADKAGTR 244
              |  |||||                    ||| |||| ||  ||| | | :
Db        919 GKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGAAGKEGAK 969


RESULT 13
US-10-995-561-912
; Sequence 912, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
```

```
;   ORGANISM: Homo sapiens
US-10-821-234-1431

  Query Match             12.4%;  Score 261;  DB 6;  Length 1366;
  Best Local Similarity   32.9%;  Pred. No. 1.3e-12;
  Matches   80;  Conservative   10;  Mismatches   89;  Indels   64;  Gaps   10;

Qy         13 PAAAPRERGSQGCGCG----GAPARAGE-GNSCLLFLGFFGLSLALHLLTLCCYLELRSE 67
              |:  | |||: |    |    | |   || ||      |  |
Db        638 PSGLPGERGAAGIPGGKGEKGEPGLRGEIGNP-----GRDG------------------- 673

Qy         68 LRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDS 127
                 |||   :|  |  | :|     |   |  |         |:  | | ||
Db        674 ---ARGAHGAVGAPGPAGATGDRGEAGAAGPAGP-------AGPRGSPGERGEVG----- 718

Qy        128 QDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGK--KAGPPGPN 185
                 | | |        :  ::  |  | |   |  ||      |    ||| |||
Db        719 -----PAGPNGFAGPAGAAGQPGAKGERGAKGPKGENGVVGPT-----GPVGAAGPAGPN 768

Qy        186 GPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPP------GPPGPPGPQGPPGLQGPSGAAD 239
              ||||| |  |  ||||: | ||  | |  |||      ||||||||  |  ||:|| |
Db        769 GPPGPAGSRGDGGPPGMTGFPGAAGRT--GPPGPSGISGPPGPPGPAGKEGLRGPRGDQG 826

Qy        240 KAG 242
                |
Db        827 PVG 829


RESULT 8
US-11-186-284-31
; Sequence 31, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/11/186,284
;  CURRENT FILING DATE:  2005-07-21
;  PRIOR APPLICATION NUMBER: US/10/301,822
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 1366
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-11-186-284-31

  Query Match             12.4%;  Score 261;  DB 7;  Length 1366;
  Best Local Similarity   32.9%;  Pred. No. 1.3e-12;
  Matches   80;  Conservative   10;  Mismatches   89;  Indels   64;  Gaps   10;

Qy         13 PAAAPRERGSQGCGCG----GAPARAGE-GNSCLLFLGFFGLSLALHLLTLCCYLELRSE 67
              |:  | |||: |    |    | |   || ||      |  |
Db        638 PSGLPGERGAAGIPGGKGEKGEPGLRGEIGNP-----GRDG------------------- 673

Qy         68 LRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDS 127
                 |||   :|  |  | :|     |   |  |         |:  | | ||
Db        674 ---ARGAHGAVGAPGPAGATGDRGEAGAAGPAGP-------AGPRGSPGERGEVG----- 718

Qy        128 QDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGK--KAGPPGPN 185
                 | | |        :  ::  |  | |   |  ||      |    ||| |||
Db        719 -----PAGPNGFAGPAGAAGQPGAKGERGAKGPKGENGVVGPT-----GPVGAAGPAGPN 768

Qy        186 GPPGPPGPPGPQGPPGIPGIPGIPGTTVMGPP------GPPGPPGPQGPPGLQGPSGAAD 239
              ||||| |  |  ||||: | ||  | |  |||      ||||||||  |  ||:|| |
Db        769 GPPGPAGSRGDGGPPGMTGFPGAAGRT--GPPGPSGISGPPGPPGPAGKEGLRGPRGDQG 826

Qy        240 KAG 242
                |
Db        827 PVG 829


RESULT 9
US-10-220-824-8
; Sequence 8, Application US/10220824
; Publication No. US20050277603A1
; GENERAL INFORMATION:
;  APPLICANT: ViroMed Limited
;  TITLE OF INVENTION: Compositions for gene therapy of rheumatoid arthritis including a
;  TITLE OF INVENTION:  gene encoding an anti-angiogenic protein or parts thereof
;  FILE REFERENCE: OPF0208/PCT
;  CURRENT APPLICATION NUMBER: US/10/220,824
;  CURRENT FILING DATE:  2002-08-30
;  PRIOR APPLICATION NUMBER: KR 2001-0000691
;  PRIOR FILING DATE: 2001-01-05
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 8
;   LENGTH: 1516
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-220-824-8

  Query Match             12.4%;  Score 261;  DB 6;  Length 1516;
  Best Local Similarity   28.0%;  Pred. No. 1.5e-12;
  Matches   71;  Conservative   19;  Mismatches   78;  Indels   86;  Gaps    7;

Qy          61 YLELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGE 120
               :|:| :|:: |:|      |  |  |  |      |  |  |      | | |:  |
Db        1081 FLQLEAEMKGEKGDRGDAGQKGERGEPGGGGFFGSSLPGPP-----GPPGPRGYP----- 1130

Qy         121 AALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAG 180
                                                  |  ||      :| : |
Db        1131 ----------------------------------------GIPGPKGESIRG-QPG 1145

Qy         181 PPGPNGPP--------GPPGPPGPQGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQ 232
               |||| |||        |||||||| |||  ||     |     :  ||||||||| ||||
Db        1146 PPGPQGPPGIGYEGRQGPPGPPGPPGPPSFPG----PHRQTISVPGPPGPPGPPGPPGTM 1201

Qy         233 GPSGAADKAGTRENQPAVVHLQGQGSAIQV--KNDLSGGVLNDWSRI------------- 277
               | |       ||:     ||   :|  | |  : :|   | | : ::
Db        1202 GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRGTD 1261

Qy         278 ----TMNPKVFKLH 287
                    : | | :||
Db        1262 NEVAALQPPVVQLH 1275


RESULT 10
US-11-186-284-33
; Sequence 33, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
```

```
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-11-186-284-28

  Query Match             12.7%;  Score 265.5;  DB 7;  Length 1464;
  Best Local Similarity   29.7%;  Pred. No. 6.3e-13;
  Matches   78;  Conservative   16;  Mismatches   80;  Indels   89;  Gaps    9;

Qy          13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
               || || ::|  |         |  |||   ||     | :      || |
Db         774 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------- 815

Qy          62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
                  ||: : |  | || :|      |   |  |      | | |       ||
Db         816 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 861

Qy         122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                  | :: : |  |          : ||
Db         862 ----------------------------------KGARGSAGPPGATGFPGAAGRVGP 885

Qy         182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
               |||:|  ||||||||      :|| |  |  | ||   :|||||||| | :| ||  ||:
Db         886 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 943

Qy         236 GAADKAGTR--ENQPAVVHLQGQ 256
               ||    | :    |  || | ||
Db         944 GAPGTPGPQGIAGQRGVVGLPGQ 966


RESULT 5
US-11-021-603-2
; Sequence 2, Application US/11021603
; Publication No. US20060003954A1
; GENERAL INFORMATION:
;  APPLICANT: Beri, Rajinder
;  TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
;  FILE REFERENCE: 06275-254US1
;  CURRENT APPLICATION NUMBER: US/11/021,603
;  CURRENT FILING DATE:  2004-12-21
;  PRIOR APPLICATION NUMBER: US/10/149,352
;  PRIOR FILING DATE: 2002-06-10
;  PRIOR APPLICATION NUMBER: PCT/GB00/04741
;  PRIOR FILING DATE: 2000-12-12
;  PRIOR APPLICATION NUMBER: GB 9929487.8
;  PRIOR FILING DATE: 1999-12-15
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
;   LENGTH: 1464
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-021-603-2

  Query Match             12.7%;  Score 265.5;  DB 7;  Length 1464;
  Best Local Similarity   29.7%;  Pred. No. 6.3e-13;
  Matches   78;  Conservative   16;  Mismatches   80;  Indels   89;  Gaps    9;

Qy          13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
               || || ::|  |         |  |||   ||     | :      || |
Db         774 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------- 815

Qy          62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
                  ||: : |  | || :|      |   |  |      | | |       ||
Db         816 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 861

Qy         122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                  | :: : |  |          : ||
Db         862 ----------------------------------KGARGSAGPPGATGFPGAAGRVGP 885

Qy         182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
               |||:|  ||||||||      :|| |  |  | ||   :|||||||| | :| ||  ||:
Db         886 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 943

Qy         236 GAADKAGTR--ENQPAVVHLQGQ 256
               ||    | :    |  || | ||
Db         944 GAPGTPGPQGIAGQRGVVGLPGQ 966


RESULT 6
US-10-821-234-1096
; Sequence 1096, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1096
;   LENGTH: 1467
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1096

  Query Match             12.7%;  Score 265.5;  DB 6;  Length 1467;
  Best Local Similarity   29.7%;  Pred. No. 6.3e-13;
  Matches   78;  Conservative   16;  Mismatches   80;  Indels   89;  Gaps    9;

Qy          13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
               || || ::|  |         |  |||   ||     | :      || |
Db         777 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------- 818

Qy          62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
                  ||: : |  | || :|      |   |  |      | | |       ||
Db         819 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 864

Qy         122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                  | :: : |  |          : ||
Db         865 ----------------------------------KGARGSAGPPGATGFPGAAGRVGP 888

Qy         182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
               |||:|  ||||||||      :|| |  |  | ||   :|||||||| | :| ||  ||:
Db         889 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 946

Qy         236 GAADKAGTR--ENQPAVVHLQGQ 256
               ||    | :    |  || | ||
Db         947 GAPGTPGPQGIAGQRGVVGLPGQ 969


RESULT 7
US-10-821-234-1431
; Sequence 1431, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1431
;   LENGTH: 1366
;   TYPE: PRT
```

```
RESULT 2
US-10-821-234-914
; Sequence 914, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 914
;   LENGTH: 1532
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-914

  Query Match             13.0%;  Score 273;  DB 6;  Length 1532;
  Best Local Similarity   31.2%;  Pred. No. 1.8e-13;
  Matches   85;  Conservative   13;  Mismatches   64;  Indels  110;  Gaps   13;

Qy         20 RGSQGC----GCGGAPARAGEGNSCLLFLGFFGLSLALHLLTLCCYLELRSELRRERGAE 75
              || ||      |  | |  :|:        |  ||:                    | :
Db        789 RGEQGLTGMPGIRGPPGPSGDP-------GKPGLT----------------------GPQ 819

Qy         76 SRLGGSGTPGTSGTLSSLGGLDPDSPITSH----------------LGQPSPKQQPLEPG 119
                |  ||||  |     |   :||                   :| | |   |   |
Db        820 GPQGLPGTPGRPGIKGEPGA--PGKIVTSEGSSMLTVPGPPGPPGAMGPPGPPGAPGPAG 877

Qy        120 EAALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKA 179
              | |      |||  |                                ||
Db        878 PAGL-----PGHQEVL----------------------------NLQGP---------P 894

Qy        180 GPPGPNGPPGP--PGPPGPQGPP--GIPGIPGIPG-------TTVMGPPGPPGPPGP--- 225
              ||||| |||||  ||||||:|||  |:|| || ||       | : ||||||||||||
Db        895 GPPGPRGPPGPSIPGPPGPRGPPGEGLPGPPGPPGSFLSNSETFLFGPPGPPGPPGPKGD 954

Qy        226 QGPPGLQGPSGAADKAG--TRENQPAVVHLQG 255
              ||||| :|  |     |  |  :    ::|||
Db        955 QGPPGPRGHQGEQGLPGFSTSGSSSFGLNLQG 986


RESULT 3
US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Qian, Xiaohong B.
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Cao, Yi-Cheng
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;  FILE REFERENCE: 785CIP4CN
;  CURRENT APPLICATION NUMBER: US/11/000,463
;  CURRENT FILING DATE:  2004-11-29
;  PRIOR APPLICATION NUMBER: 10/291,265
;  PRIOR FILING DATE: 2002-11-08
;  PRIOR APPLICATION NUMBER: PCT/US01/02623
;  PRIOR FILING DATE: 2001-01-25
;  PRIOR APPLICATION NUMBER: 09/922,279
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: 09/491,404
;  PRIOR FILING DATE: 2000-01-25
;  PRIOR APPLICATION NUMBER: 09/617,746
;  PRIOR FILING DATE: 2000-07-17
;  PRIOR APPLICATION NUMBER: 09/631,451
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 09/633,870
;  PRIOR FILING DATE: 2000-09-15
;  NUMBER OF SEQ ID NOS: 944
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 243
;   LENGTH: 1464
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-000-463-243

  Query Match             12.7%;  Score 265.5;  DB 7;  Length 1464;
  Best Local Similarity   29.7%;  Pred. No. 6.3e-13;
  Matches   78;  Conservative   16;  Mismatches   80;  Indels   89;  Gaps    9;

Qy         13 PAAAPRERGSQG-------CGCGGAPARAGE----GNSCLLFLGFFGLSLALHLLTLCCY 61
              || || ::|  |         |  |||   ||     | :     || |
Db        774 PAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPA-----GFAG------------- 815

Qy         62 LELRSELRRERGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEA 121
                    ||: : |  | || :|      |   |  |      | | |      ||
Db        816 ---------PPGADGQPGAKGEPGDAGAKGDAGPPGPAGP----AGPPGPIGNVGAPGA- 861

Qy        122 ALHSDSQDGHQMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGP 181
                                                  | :: : |  |          : ||
Db        862 ----------------------------------KGARGSAGPPGATGFPGAAGRVGP 885

Qy        182 PGPNGPPGPPGPPGP------QGPPGIPGIPGIPGTTVMGPPGPPGPPGPQGPPGLQGPS 235
              |||:|  ||||||||      :|| |  |  | ||   :||||||| | :| ||  ||:
Db        886 PGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGE--VGPPGPPGPAGEKGSPGADGPA 943

Qy        236 GAADKAGTR--ENQPAVVHLQGQ 256
              ||    | :    |  || | ||
Db        944 GAPGTPGPQGIAGQRGVVGLPGQ 966


RESULT 4
US-11-186-284-28
; Sequence 28, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/11/186,284
;  CURRENT FILING DATE:  2005-07-21
;  PRIOR APPLICATION NUMBER: US/10/301,822
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
;   LENGTH: 1464
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2006, 16:15:28 ; Search time 31 Seconds
(without alignments)
119.245 Million cell updates/sec

Title: US-09-729-658C-2
Perfect score: 2098
Sequence: 1 MGYPEVERRELLPAAAPRER.....NMSKHTTFFGAIRLGEAPAS 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 67062 seqs, 9454214 residues

Total number of hits satisfying chosen parameters: 67062

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	280.5	13.4	1874	6	US-10-821-234-1182	Sequence 1182, Ap
2	273	13.0	1532	6	US-10-821-234-914	Sequence 914, App
3	265.5	12.7	1464	7	US-11-000-463-243	Sequence 243, App
4	265.5	12.7	1464	7	US-11-186-284-28	Sequence 28, Appl
5	265.5	12.7	1464	7	US-11-021-603-2	Sequence 2, Appli
6	265.5	12.7	1467	6	US-10-821-234-1096	Sequence 1096, Ap
7	261	12.4	1366	6	US-10-821-234-1431	Sequence 1431, Ap
8	261	12.4	1366	7	US-11-186-284-31	Sequence 31, Appl
9	261	12.4	1516	6	US-10-220-824-8	Sequence 8, Appli
10	256	12.2	1466	7	US-11-186-284-33	Sequence 33, Appl
11	254	12.1	1767	6	US-10-995-561-911	Sequence 911, App
12	254	12.1	1767	6	US-10-995-561-914	Sequence 914, App
13	254	12.1	1806	6	US-10-995-561-912	Sequence 912, App
14	254	12.1	1806	6	US-10-995-561-915	Sequence 915, App
15	254	12.1	1818	6	US-10-995-561-910	Sequence 910, App
16	254	12.1	1818	6	US-10-995-561-913	Sequence 913, App
17	253	12.1	744	7	US-11-186-284-37	Sequence 37, Appl
18	252	12.0	744	7	US-11-186-284-39	Sequence 39, Appl
19	246	11.7	1736	7	US-11-124-368A-329	Sequence 329, App
20	245	11.7	1166	6	US-10-821-234-964	Sequence 964, App
21	243	11.6	1496	7	US-11-186-284-35	Sequence 35, Appl
22	236	11.2	3063	7	US-11-186-284-26	Sequence 26, Appl
23	233	11.1	828	6	US-10-995-561-983	Sequence 983, App
24	233	11.1	918	6	US-10-995-561-981	Sequence 981, App
25	233	11.1	1019	6	US-10-995-561-982	Sequence 982, App
26	221.5	10.6	551	7	US-11-096-070-12	Sequence 12, Appl
27	220.5	10.5	520	6	US-10-995-561-532	Sequence 532, App
28	209	10.0	549	7	US-11-096-070-8	Sequence 8, Appli
29	203.5	9.7	287	7	US-11-174-150-30	Sequence 30, Appl
30	197	9.4	549	7	US-11-096-070-2	Sequence 2, Appli
31	192	9.2	549	7	US-11-096-070-6	Sequence 6, Appli
32	191	9.1	427	7	US-11-096-070-10	Sequence 10, Appl
33	189	9.0	531	7	US-11-096-070-4	Sequence 4, Appli
34	182	8.7	253	6	US-10-821-234-1438	Sequence 1438, Ap
35	176.5	8.4	406	6	US-10-131-826A-82	Sequence 82, Appl
36	168	8.0	120	7	US-11-096-070-35	Sequence 35, Appl
37	167	8.0	356	7	US-11-143-986-8	Sequence 8, Appli
38	167	8.0	356	7	US-11-143-986-9	Sequence 9, Appli
39	166.5	7.9	117	7	US-11-096-070-34	Sequence 34, Appl
40	162.5	7.7	288	7	US-11-135-855-30	Sequence 30, Appl
41	157.5	7.5	451	6	US-10-995-561-886	Sequence 886, App
42	156	7.4	303	7	US-11-135-855-31	Sequence 31, Appl
43	152.5	7.3	498	7	US-11-150-845-40	Sequence 40, Appl
44	148	7.1	388	6	US-10-995-561-887	Sequence 887, App
45	148	7.1	501	7	US-11-134-563-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-10-821-234-1182
; Sequence 1182, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1182
; LENGTH: 1874
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1182

Query Match 13.4%; Score 280.5; DB 6; Length 1874;
Best Local Similarity 30.1%; Pred. No. 5.9e-14;
Matches 83; Conservative 9; Mismatches 69; Indels 115; Gaps 10;

```
Qy          13 PAAAPRERGSQGC-GCGGAPARAGEGNSCLLFLGPFGLSLALHLLTLCCYLELRSELRRE 71
               |   | |:|  |  |  | | | |  |    :|| |
Db         893 PLGPPGEKGKLGVPGLPGYPGRQGPKGS----IGFPGFP---------------------- 927

Qy          72 RGAESRLGGSGTPGTSGTLSSLGGLDPDSPITSHLGQPSPKQQPLEPGEAALHSDSQDGH 131
                 ||    || |||                          |:| |: |    |
Db         928 -GANGEKGGRGTP----------------------GKPGPRGQRGPTG------------ 952

Qy         132 QMALLNFFFPDEKPYSEEESRRVRRNKRSKSNEGADGPVKNKKKGKKAGPP---GPNGPP 188
                           |  |   | :      | | | |||          ||||   |||||
Db         953 ------------PRGERGPRGITGKPGPKGNSGGDGP---------AGPPGERGPNGPQ 990

Qy         189 GPPGPPGPQGPPGIPGIPGIPG-------TTVMGPPGPPGP------------------- 222
               || | |||:|||| ||  |:||       |   |  |||||
Db         991 GPTGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGER 1050

Qy         223 --PGPQGPPGLQGPSGAADKAGTRENQPAVVHLQGQ 256
                 ||| |||| ||  | | | ||:   |    | |:
Db        1051 GHPGPPGPPGEQGLPGLAGKEGTK-GDPGPAGLPGK 1085
```